

**THE ACCURACY OF WRIST-WORN ACTIVITY MONITORS AT MEASURING
ENERGY EXPENDITURE DURING DIFFERENT FORMS OF PHYSICAL ACTIVITY:
A PILOT STUDY**

RUBIN POONI

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

GRADUATE PROGRAM IN KINESIOLOGY AND HEALTH SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

July 2019

© Rubin Pooni, 2019

ABSTRACT

The accuracy of activity monitors are critical to the proper functioning of artificial pancreas (AP) systems. The objective of this thesis was to assess the accuracy of activity monitors at measuring energy expenditure (EE) during physical activity. In two separate studies, the Fitbit Charge 2 (FC2), Garmin vivosmart HR+ (GVHR+), Fitbit Ionic (FI), and Garmin vivosmart 3 (GV3) were assessed during 4-5 types of physical activity in 15-20 adults. In Study A, the GVHR+ displayed significantly less negative bias in EE than the FC2 (FC2: $-19.3 \pm 28.9\%$, GVHR+: $-1.6 \pm 30.6\%$, $P < 0.001$). In Study B, the GV3 displayed less positive bias in EE than the FI (FI: $15.7 \pm 36.0\%$, GV3: $7.0 \pm 36.3\%$). The GV3 significantly overestimated EE in T1D subjects compared to healthy controls. None of the activity monitors tested are ready to be incorporated into an AP system, but the GV3 is approaching the accuracy required.

TABLE OF CONTENTS

Abstract.....	ii
Table of Contents	iii
1. Introduction	1
1.1. Type 1 Diabetes	1
1.2. Activity Monitors	4
1.3. Artificial Pancreas	5
2. Review of the Scientific Literature	7
2.1. Accuracy of Activity Monitors during Physical Activity	7
2.1.1. Walking.....	7
2.1.2. Running	13
2.1.3. Elliptical Exercise	17
2.1.4. Cycling.....	18
2.1.5. Resistance Exercise	22
2.1.6. Sports and Sports-Related Exercise	24
2.2. Accuracy of Activity Monitors during Daily Living and Sedentary Activities ..	26
2.2.1. Activities of Daily Living	26
2.2.2. Sedentary Activities	29
2.3. Accuracy of Activity Monitors in Individuals with Diabetes	31
2.4. Summary	33
3. Study Overview, Objectives & Hypotheses.....	35
3.1. Study Overview	35
3.2. Primary Objectives	35
3.3. Secondary Objectives	36
3.4. Hypotheses	36
4. Study A: Pre-Pilot Study	38
4.1. Methods.....	38
4.1.1. Participants.....	38
4.1.2. Study Protocol.....	38
4.1.2.1. Exercise Visit #1	40
4.1.2.2. Exercise Visit #2.....	42
4.1.3. Devices.....	44
4.1.3.1. Research-Grade Devices	44
4.1.3.2. Consumer-Grade Devices	45
4.1.4. Statistical Analysis	46
4.2. Results.....	48

4.2.1.	Participants.....	48
4.2.2.	Data Loss	48
4.2.3.	Energy Expenditure Accuracy	49
4.2.4.	METs and VO ₂ Accuracy	52
5.	Study B: Pilot Study	54
5.1.	Methods.....	54
5.1.1.	Participants.....	54
5.1.2.	Study Protocol.....	54
5.1.2.1.	Exercise Visit #1	56
5.1.2.2.	Exercise Visit #2.....	59
5.1.3.	Devices.....	60
5.1.3.1.	Research-Grade Devices	60
5.1.3.2.	Consumer-Grade Devices	61
5.1.4.	Glucose	62
5.1.5.	Statistical Analysis	64
5.2.	Results.....	66
5.2.1.	Participants.....	66
5.2.2.	Data Loss	67
5.2.3.	Energy Expenditure Accuracy	68
5.2.4.	Glucose	72
6.	Discussion.....	75
6.1.	Principal Findings.....	75
6.2.	Strengths	81
6.3.	Limitations	82
6.4.	Future Directions.....	83
6.5.	Conclusions.....	85
7.	References.....	87
8.	Abbreviations	101
9.	Tables & Figures	103
9.1.	Summary of the Scientific Literature	103
	Table 1. Summary of Literature on Accuracy of Activity Monitors for Walking.....	103
	Table 2. Summary of Literature on Accuracy of Activity Monitors for Running	107
	Table 3. Summary of Literature on Accuracy of Activity Monitors for Elliptical Exercise	111

Table 4. Summary of Literature on Accuracy of Activity Monitors for Cycling.....	112
Table 5. Summary of Literature on Accuracy of Activity Monitors for Resistance Exercise	114
Table 6. Summary of Literature on Accuracy of Activity Monitors for Sports and Sports-Related Exercise	115
Table 7. Summary of Literature on Accuracy of Activity Monitors for Activities of Daily Living	116
Table 8. Summary of Literature on Accuracy of Activity Monitors for Sedentary Activities	117
Table 9. Summary of Literature on Accuracy of Activity Monitors in Individuals with Diabetes	118
9.2. Study A	119
Table 10. Participant Characteristics	119
Table 11. Energy Expenditure.....	120
Figure 1. Total Energy Expenditure	122
Figure 2. Correlation Plots for Total Energy Expenditure.....	123
Figure 3. Absolute Percent Error in Energy Expenditure	124
Figure 4. Relative Percent Error in Energy Expenditure	124
Figure 5. Bland-Altman Plots for Energy Expenditure.....	125
Figure 6. Relative Percent Error in Energy Expenditure for Treadmill and Cycle Ergometer Exercise	125
Figure 7. Correlation Plots for Energy Expenditure.....	126
9.3. Study B	127
Table 12. Participant Characteristics	127
Table 13. Energy Expenditure.....	128
Figure 8. Total Energy Expenditure	129
Figure 9. Correlation Plots for Total Energy Expenditure.....	131
Figure 10. Absolute Percent Error in Energy Expenditure	132
Figure 11. Relative Percent Error in Energy Expenditure	132
Figure 12. Bland-Altman Plots for Energy Expenditure.....	133
Figure 13. Relative Percent Error in Energy Expenditure for Treadmill and Cycle Ergometer Exercise	133
Figure 14. Relative Percent Error in Energy Expenditure in T1D and Control Subjects.....	134
Figure 15. Blood Glucose.....	134
Figure 16. Change in Blood Glucose	135
Figure 17. Percent Time in Hypoglycemia, Euglycemia, and Hyperglycemia	136
Figure 18. Interstitial Glucose	137

10. Appendices.....	138
10.1. Study A.....	138
Appendix A. Activity Modes.....	138
Appendix B. Energy Expenditure for Individual Subjects during Cycling Exercise.....	139
Appendix C. Relative Percent Error in Energy Expenditure for Dominant and Non-Dominant Hand Wear	140
Appendix D. Correlation Plots for Relative Percent Error and Anthropometrics	144
Appendix E. Relative Percent Error in METs and VO2	145
10.2. Study B.....	146
Appendix F. Activity Modes.....	146
Appendix G. Relative Percent Error in Energy Expenditure on Garmin vívosmart 3.....	147
Appendix H. Fitzpatrick Skin Tone Score and Body Fat Percentage in T1D and Control Subjects.....	148
Appendix I. Relative Percent Error in Energy Expenditure for Dominant and Non-Dominant Hand Wear	148
Appendix J. Correlation Plots for Relative Percent Error and Anthropometrics	152

1. INTRODUCTION

1.1. TYPE 1 DIABETES

Type 1 diabetes (T1D) is an autoimmune disease characterized by the destruction of insulin-producing β -cells in the pancreas, resulting in total insulin deficiency and hyperglycemia in the body (1). Consequently, individuals with T1D are required to administer insulin exogenously to manage blood glucose (BG) to try and maintain it within euglycemic range (3.9-10.0 mmol/L) (1–3). Over time, T1D can lead to long-term complications, such as cardiovascular disease (CVD), retinopathy, nephropathy, and neuropathy (4).

The exact number of individuals with T1D around the world is unknown, but in the United States, it is estimated that up to 3 million individuals live with T1D (1). Approximately 64 000 new cases of T1D develop annually in the United States in individuals between 0 to 64 years of age (5). Although no formal assessment has been conducted in Canada, it is estimated that more than 300 000 Canadians are living with T1D (6).

Individuals with T1D are prescribed to either multiple daily insulin injections (MDI) or continuous subcutaneous insulin infusion (CSII) to help manage glycemic control (7). With MDI, individuals administer a long-acting insulin with a syringe as a basal insulin and administer a rapid-acting insulin with a syringe prior to meals. The amount of rapid-acting insulin administered is based on the grams of carbohydrates consumed (4). CSII involves the use of a small, portable device (i.e. an insulin pump) that infuses rapid-acting insulin at a slow basal rate for 24 hours a day through a cannula inserted into

subcutaneous tissue. Individuals administer larger doses (i.e. boluses) of insulin at mealtime using the insulin pump based on the grams of carbohydrates consumed (8).

Individuals with T1D use self-monitoring of blood glucose (SMBG) with a hand-held glucose meter, or continuous glucose monitoring (CGM) with an implanted glucose sensor, to monitor their glucose control (7). SMBG typically involves piercing a finger with a lancet device to acquire a small drop of blood. The drop of blood is applied to a reagent strip that is inserted into a hand-held glucose meter (i.e. glucometer), which will measure glucose from the strip and display it on a screen (9). Most individuals using SMBG may test up to 6 to 10 times a day, but this can vary based on individual needs and motivations (7). CGM reduces the frequency of SMBG by providing measurements approximately every 5 minutes from a sensor inserted into subcutaneous tissue that measures the amount of glucose in interstitial fluid. The values are transmitted from the sensor to a data storage device or a sensor-associated app on the user's smartphone via Bluetooth (10). SMBG and CGM allow individuals to monitor their BG and make necessary adjustments to insulin, food intake, or exercise to maintain glycemic control (7).

Although SMBG and CGM can be effective tools for daily self-management of T1D, Diabetes Canada and the American Diabetes Association (ADA) recommend that overall glycemic management be assessed using glycated hemoglobin (HbA_{1c}). HbA_{1c} reflects an individual's average glycemia over ~3 months. Diabetes Canada and the ADA recommend that adults with T1D maintain a HbA_{1c} of $\leq 7\%$ to reduce the risk of microvascular and macrovascular complications (11,12). In a longitudinal study of 1441 people with T1D, participants on intensive insulin therapy (≥ 3 daily insulin injections with

MDI or CSII) with a mean HbA_{1c} of 7.2% had a 63% ($P \leq 0.002$), 39-54% (39%: $P \leq 0.002$; 54%: $P < 0.04$), and 60% ($P \leq 0.002$) decrease in the risk of development and progression of retinopathy, nephropathy, and neuropathy, respectively, compared to participants on conventional therapy (1-2 daily insulin injections) with a mean HbA_{1c} of 9.1% (13). At a 30-year follow-up, the individuals on intensive insulin therapy (≥ 3 daily insulin injections with MDI or CSII) had a mean HbA_{1c} of 7.8% and displayed a decrease in the risk of CVD by 30% ($P = 0.016$) and the incidence of major cardiovascular events (nonfatal myocardial infarction, stroke, or cardiovascular death) by 32% ($P = 0.07$) compared to the conventional therapy group (1-2 daily insulin injections), who had a mean HbA_{1c} of 8.2%. The lower HbA_{1c} levels in the intensive insulin therapy group were determined to statistically account for all of the observed treatment effects on CVD risk in this group (14).

Diabetes Canada and the ADA also recommend that adults with diabetes (type 1 and type 2) engage in ≥ 150 minutes of moderate- (64-76% of a person's maximum heart rate (HR)) to vigorous-intensity ($> 76\%$ of a person's maximum HR) aerobic exercise over at least 3 non-consecutive days of the week. Adults who are willing and able to can also perform interval training. Interval training consists of alternating between short periods (30 seconds to 3 minutes) of higher- and lower-intensity exercise. In addition to aerobic exercise, adults with diabetes should perform resistance exercise on ≥ 2 non-consecutive days of the week (15,16). However, most adults living with T1D do not meet these guidelines (17). Performing regular exercise training at 50-90% of peak oxygen consumption (VO_2 peak) at least twice a week for a minimum of 8

weeks has been demonstrated to decrease HbA_{1c} by 0.78% ($P<0.0001$) and increase maximal oxygen consumption by 3.45 mL/kg/min ($P=0.02$) in individuals with T1D (18).

1.2. ACTIVITY MONITORS

Physical activity monitors have become increasingly popular in recent years, with 423 new devices released from 2011 to 2017 (19). Activity monitors are small devices that are commonly worn on the wrist. They provide the user with real-time feedback on different physiological measurements and daily activities, such as HR, energy expenditure (EE), step count, and sleep duration (20). Through synchronization with a computer- or mobile-based app, users are able to receive more detailed feedback, such as summary data and progress towards personalized goals, as well as the ability to connect to social media and other health and fitness apps (20,21). Activity monitors can be effective tools for increasing physical activity (22).

These devices can use one or more different types of sensors, including pedometers, accelerometers, gyroscopes, magnetometers, barometers, altimeters, global positioning systems (GPS), and photoplethysmography (PPG), to measure different metrics (19). Activity-specific algorithms incorporated into the device software analyze the pattern of signals from sensors to calculate EE and other physiological and activity outputs (23). Although the physical sensors in most devices are similar, the algorithms being used are unique to the manufacturer and are often proprietary (19).

1.3. ARTIFICIAL PANCREAS

Artificial pancreas (AP) systems are being developed and tested to help individuals with T1D maintain euglycemia by automatically adjusting to variations caused by daily activities and disturbances. The four major sources of variations in BG concentrations are food, exercise, sleep, and stress (24).

All AP systems consist of three basic components: a glucose sensor (e.g. CGM), insulin infusion mechanisms (e.g. insulin pumps), and decision-making algorithms that are used to calculate insulin infusion needs. Some researchers have proposed that physical activity monitors may help to better inform decision-making algorithms, since physical activity changes minute-by-minute insulin needs (25). To date, only one AP model (Medtronic MiniMed 670G Hybrid Closed-Loop System) has been approved for clinical use, but several others are in development. In models under development, glucose and physiological data gathered by sensors are inputted into a feedback control algorithm, which makes the necessary adjustments to an insulin pump to maintain euglycemia (24). In one study conducted by Turksoy et al. (26), the use of an AP system in individuals with T1D on a physically active day (20 minutes of treadmill running at 70-97% of their age-predicted maximum HR) resulted in a significant decrease in mild hypoglycemia (3.0-3.9 mmol/L) compared to individuals without an AP system ($P<0.01$). Zero percent (0/9) of participants using the AP system experienced severe hypoglycemia (<3.0 mmol/L) compared to 82% (9/11) of participants who were not using the AP system ($P<0.01$). In addition, there was a significant difference in the time spent in euglycemic range between the AP group and the group without AP (AP: 58%; No AP: 54%; $P<0.01$).

For any medical device, including AP systems, the accuracy of the sensors (i.e. CGM, activity monitors, etc.) being used are critical to the proper functioning of the system, since they affect the decisions being made by the medical device and patient outcome. Therefore, it is important that the accuracy of these sensors be assessed in different conditions and in the populations by which they will be used. This thesis focuses on the accuracy and utility of physical activity monitors for the purpose of developing an AP system that takes into account various metrics of physical activity (e.g. HR, EE, etc.) into the control algorithm.

2. REVIEW OF THE SCIENTIFIC LITERATURE

2.1. ACCURACY OF ACTIVITY MONITORS DURING PHYSICAL ACTIVITY

The scientific literature examining the accuracy of consumer-grade activity monitors at measuring EE during physical activity can be divided into six different types of activity: walking, running, elliptical exercise, cycling, resistance exercise, and sports and sports-related exercise.

2.1.1. Walking

Twenty-four studies have examined the accuracy of EE measurements from consumer-grade activity monitors for walking (27–50). Some of these studies have examined accuracy during a combination of activities, but did not report the accuracy of the devices tested during individual activities, such as walking (28,30–32,40,41). However, many others have reported their findings for accuracy of activity monitors during walking (Table 1). Accuracy at a variety of different walking speeds has been assessed, with some studies examining accuracy during two or three different walking speeds (31,34–36,38,40,44–48), while others have observed accuracy at different treadmill inclines (42,49). The majority of these studies have been conducted for treadmill walking, while a few studies have also assessed accuracy during overground walking (27,50). In addition, two studies have assessed the accuracy of activity monitors for walking with weighted bags (38,47).

Gusmer et al. (33) had 32 healthy adults perform 30 minutes of slow walking (avg. speed: 2.5 ± 0.3 mph) and 30 minutes of brisk walking (avg. speed: 3.0 ± 0.4 mph) at self-selected walking speeds, while wearing the Fitbit Ultra (Fitbit, Inc., California,

USA) at the hip. The Fitbit Ultra underestimated EE during both slow walking ($P>0.05$) and brisk walking ($P<0.05$) compared to indirect calorimetry. The Fitbit Ultra displayed a moderate relationship with the indirect calorimeter during slow walking ($r=0.69$, $P<0.001$) and a strong relationship with the indirect calorimeter during brisk walking ($r=0.94$, $P<0.001$). The Fitbit Ultra also displayed wider 95% limits of agreement (LoA) during slow walking compared to brisk walking. This study demonstrated greater error and underestimation in EE measurements by the Fitbit Ultra at faster walking speeds (33).

Wahl et al. (39) assessed the accuracy of 10 activity monitors during two different walking speeds. Twenty healthy adults walked at 2.7 mph and 4.3 mph on a treadmill for 5 minutes at each speed, while wearing various activity monitors: Beurer AS80 (Beurer GmbH, Ulm, Germany), Fitbit Charge (Fitbit, Inc., California, USA), Fitbit Charge HR (Fitbit, Inc., California, USA), Garmin Forerunner 920XT (Garmin International, Inc., Kansas, USA), Garmin vívoactive (Garmin International, Inc., Kansas, USA), Garmin vívofit (Garmin International, Inc., Kansas, USA), Garmin vívosmart (Garmin International, Inc., Kansas, USA), Polar Loop (Polar Electro, Kempele, Finland), and Withings Pulse O_x (Withings, Issy-les-Moulineaux, France). One of each device was worn on the wrist, while an additional Withings Pulse O_x monitor was worn at the hip. All devices overestimated EE while walking at 2.7 mph by 1-83%, except for the Garmin Forerunner 920XT and Withings Pulse O_x worn at the hip and at the wrist, which underestimated by 27%, 11%, and 10%, respectively. All devices overestimated EE while walking at 4.3 mph by 17-54%, except for the Beurer AS80 and Garmin Forerunner 920XT, which underestimated by 19% and 17%, respectively. For

the 2.7 mph walking speed, the Garmin vívofit (1%), Garmin vívosmart (3%), Garmin vívoactive (4%), and Withings Pulse at the hip (-10%) were the only devices to achieve a mean relative percentage error (%RE) of $\leq 10\%$. No devices achieved this for the 4.3 mph walking speed. Similar to the Gusmer et al. (33) study, Wahl et al. observed greater error in EE at faster walking speeds in 6 out of the 10 consumer-grade activity monitors tested (39).

Price et al. (48) had subjects walk at three different speeds for 4 minutes each. Fourteen healthy adults walked at 1.6, 2.8, and 4.0 mph on a treadmill, while wearing the Fitbit One (Fitbit, Inc., California, USA) at the hip and the Garmin vívofit and the Jawbone UP (Jawbone, California, USA) on the wrist. The Fitbit One displayed a mean overestimation in EE of 2.2 kcal/min across all walking speeds compared to indirect calorimetry, while the Garmin vívofit and Jawbone UP displayed an underestimation in EE of 1.7 kcal/min and 2.1 kcal, respectively. Bias was greater as walking intensity increased for the Fitbit One and Garmin vívofit, but could not be analyzed for the Jawbone UP. This study also demonstrated increased error by activity monitors at faster walking speeds (48).

While most studies examine the accuracy of activity monitors at different walking speeds on level ground, Alsubheen et al. (42) examined accuracy in EE for walking at different inclines. Thirteen healthy adults completed self-paced walking (ranging from 2.5 to 4.5 mph) at three different inclines (0%, 5%, and 10%), while wearing the Garmin vívofit on the wrist. The Garmin vívofit significantly underestimated EE for walking at all inclines by 30% compared to indirect calorimetry ($P < 0.05$). The magnitude of underestimation displayed by the Garmin vívofit decreased as exercise intensity

increased, with the device underestimating by 45%, 30%, and 18% at 0%, 5%, and 10% inclinations, respectively. While other studies have demonstrated decreased accuracy at faster walking speeds, Alsubheen et al. were able to demonstrate increased accuracy in EE measurements for self-paced walking at higher inclines (42).

Noah et al. (49) also assessed activity monitor accuracy at different inclines. Twenty-three healthy adults walked at 3.5 mph for 6-minute bouts at a 0% and 5% incline on a treadmill, while wearing the Fitbit (Fitbit, Inc., California, USA) and the newer activity monitor, the Fitbit Ultra, at the hip. Both devices significantly underestimated EE at both inclines compared to indirect calorimetry ($P < 0.006$). Both devices displayed greater %RE at measuring EE for inclined walking ($\geq 40\%$) than for level walking. A strong intraclass correlation coefficient (ICC) with the indirect calorimeter was displayed by the Fitbit ($r = 0.70-0.72$) and Fitbit Ultra ($r = 0.81-0.83$) while walking on a 0% and 5% incline. The Fitbit Ultra displayed higher EE values and higher ICC than the Fitbit. As opposed to the Alsubheen et al. study (42), this study displayed greater error in EE measurements at higher inclines. In addition, this study was able to demonstrate improved accuracy in EE measurements by the Fitbit Ultra compared to the Fitbit. The improvements in EE accuracy may have been due to hardware upgrades, as well as improvements made to the proprietary EE algorithms by the manufacturer (49).

Diaz et al. (37) examined the accuracy of EE measurements by a wrist-worn and hip-worn activity monitor. Twenty-three healthy adults walked at 1.9, 3.0, and 4.0 mph on a treadmill for 6 minutes each, while wearing the Fitbit Flex (Fitbit, Inc., California, USA) at the wrist and the Fitbit One at the hip. The Fitbit Flex underestimated EE for

walking at 1.9 mph, while overestimating EE for walking at 3.0 and 4.0 mph compared to indirect calorimetry. The Fitbit One underestimated EE for walking at 1.9 and 3.0 mph and overestimated EE at 4.0 mph. The greatest difference in EE measurements occurred for walking at 3.0 and 4.0 mph by the Fitbit Flex, which overestimated EE by 52% and 33%, respectively. Between the two devices, the Fitbit One (hip-worn) displayed greater accuracy than the Fitbit Flex (wrist-worn) in EE measurements (37).

Diaz et al. (36) conducted a follow-up study using the same exercise protocol, but in addition to having subjects wear the Fitbit Flex on the wrist and Fitbit One at the hip, subjects also wore an additional Fitbit One on the upper torso. Thirteen healthy females walked at three speeds (1.9, 3.0, and 4.0 mph) on a treadmill for 6 minutes each, while wearing the three activity monitors. The Fitbit Flex overestimated EE, while the Fitbit One at the hip and upper torso underestimated EE at 1.9 mph compared to indirect calorimetry. All three devices overestimated EE at 3.0 and 4.0 mph. The Fitbit One at the hip displayed the greatest accuracy in EE measurements during all 3 walking speeds compared to indirect calorimetry, with a %RE in EE of -8%, 13%, and 12% at 1.9, 3.0, and 4.0 mph, respectively. Meanwhile, the Fitbit Flex at the wrist displayed the least accuracy in EE, with a %RE of 83%, 68%, and 29% at 1.9, 3.0, and 4.0 mph, respectively. With this study, Diaz et al. demonstrated that although the Fitbit One worn at the upper torso was less accurate at measuring EE than wearing it at the hip, it was still more accurate than measuring EE at the wrist by the Fitbit Flex for walking at different speeds (36).

Sasaki et al. (47) had subjects walk at two different speeds on a 5% incline, as well as walk on level ground with weighted bags. Twenty healthy adults walked at 3.0

mph and 4.0 mph on a 5% incline and walked while carrying weighted bags (4 lb each) for 6-minute bouts, while wearing the Fitbit at the hip. The Fitbit significantly underestimated EE by 22% and ~35% while walking at 3.0 mph and 4.0 mph, and significantly overestimated EE by ~40% while carrying weighted bags ($P<0.05$). Since the Fitbit was worn at the hip, Sasaki et al. hypothesized that subjects may have exacerbated hip movements while carrying the weighted bags in an effort to maintain stability while walking, which may have caused the Fitbit to overestimate EE for weighted walking (47).

Studies examining the accuracy of consumer-grade activity monitors for walking at 1.5-4.5 mph have determined most activity monitors to display poor accuracy at measuring EE (27,31,32,34–38,40–43,46–48), with many activity monitors displaying a %RE of >10% (35–37,42) and a mean absolute percentage error (MAPE) of >10% (34,38,43,47,50). Activity monitors have been demonstrated to display greater inaccuracy at faster walking speeds (28,33,48). Conflicting results have been observed in accuracy for walking at different inclines, with Alsubheen et al. (42) displaying greater accuracy at higher inclines and Noah et al. (49) displaying decreased accuracy at higher inclines. In addition, Diaz et al. demonstrated in two separate studies that a hip-worn activity monitor was more accurate than a wrist-worn activity monitor at measuring EE for walking (36,37). Sasaki et al. observed greater %RE and an overestimation in EE for walking with weighted bags, which may have been due to increased hip movement required for stability during weighted walking (47).

2.1.2. Running

Twenty-three studies have examined the accuracy of EE measurements from consumer-grade activity monitors for running (27,28,30–32,34–41,43–52). Some of these studies have examined accuracy during a combination of activities, but did not report the accuracy of the devices tested during individual activities, such as running (28,30–32,40,41). However, many others have reported their findings for accuracy of activity monitors during running (Table 2). Accuracy at various running speeds has been assessed, with some studies only examining accuracy at one speed (34,36–38,43,47,49) and others examining multiple running speeds (35,39,44,46,48,51). The majority of these studies have been conducted for treadmill running, while a few studies have also assessed accuracy during overground running (28,39,50).

Noah et al. (49) had 23 healthy adults run for 6 minutes at 5.5 mph, while wearing the Fitbit and the newer activity monitor, the Fitbit Ultra, at the hip. Both devices significantly underestimated EE for running compared to indirect calorimetry ($P<0.006$). A strong ICC with the indirect calorimeter was displayed by the Fitbit Ultra ($r=0.87$) and a moderate ICC was displayed by the Fitbit ($r=0.56$). The Fitbit Ultra displayed higher EE values and higher ICC than the Fitbit. This study was able to demonstrate improved accuracy in EE measurements by the Fitbit Ultra compared to the Fitbit for running. The improvements in EE accuracy may have been due to hardware upgrades, as well as improvements made to the proprietary EE algorithms by the manufacturer (49).

Price et al. (48) had subjects run on a treadmill at three different speeds. Fourteen healthy adults ran at 5.0, 6.2, and 7.5 mph for 4 minutes at each speed, while wearing the Fitbit One at the hip and the Garmin vívofit and the Jawbone UP at the

wrist. The Fitbit One and Jawbone UP displayed a mean overestimation in EE of 3.6 kcal/min and 20.6 kcal, respectively, across all running speeds compared to indirect calorimetry, while the Garmin vívofit displayed an underestimation in EE of 1.5 kcal/min. Bias was greater as running intensity increased for the Fitbit One, but not the Garmin vívofit. The Jawbone UP displayed greater bias during running (20.6 kcal) compared to walking (-2.1 kcal), which also indicated greater error at higher exercise intensities. This study was essentially able to demonstrate increased error in EE measurements by two out three activity monitors tested at faster running speeds (48).

Kendall et al. (52) had 50 subjects perform a graded maximal test on a treadmill, while wearing the Basis B1 (Intel Corporation, California, USA) and Fitbit Flex on the wrist, the Jawbone UP and OMRON HJ-321 (OMRON Corporation, Kyoto, Japan) on the hip, and the Polar FT7 (Polar Electro, Kempele, Finland) at the chest. Subjects began running at a self-selected speed and 0% incline and then incline was increased by 2% every 2 minutes until exhaustion. The Basis B1, Fitbit Flex, Jawbone UP24, and Polar FT7 all significantly overestimated EE, while the OMRON HJ-321 significantly underestimated EE during the graded maximal test compared to indirect calorimetry ($P<0.05$). The Polar FT7 ($r=0.74$) and Jawbone UP ($r=0.75$) displayed a moderate ICC with indirect calorimetry, while the Basis B1 ($r=0.80$), Fitbit Flex ($r=0.81$) and OMRON HJ-321 ($r=0.83$) displayed a strong ICC for the maximal test. When the ICC of devices worn by subjects that ran at slower speeds and lower inclines was compared to those who ran at higher speeds and reached a higher incline, the activity monitors displayed lower ICC at higher speeds and inclines compared to slower speeds and lower inclines, indicating decreased EE measurement accuracy at higher exercise intensities (52).

Roos et al. (51) had 20 healthy adults run at 30%, 50%, 70%, 90%, and 110% of VO_2 peak on a treadmill, while wearing the Garmin Forerunner 920XT, Polar V800 (Polar Electro, Kempele, Finland), and Suunto Ambit2 (Suunto, Vantaa, Finland) on the wrists. Subjects ran for 10-minute bouts at 30%, 50%, and 70% of VO_2 peak (low- to moderate-intensity) and ran for 90 seconds at 90% and 110% of VO_2 peak (high-intensity). The Garmin Forerunner 920XT underestimated EE ($P=0.01$) and the Suunto Ambit2 significantly overestimated EE ($P=0.002$) for running at 30% of VO_2 peak. The Suunto Ambit2 also significantly overestimated EE for running at 50% of VO_2 peak ($P=0.003$). For anaerobic running at 90% and 110% of VO_2 peak, all devices significantly underestimated EE compared to indirect calorimetry ($P<0.001$). Similar to the findings of previously discussed studies, this underestimation in EE increased as running intensity increased (51).

Dondzila and Garner (35) examined the accuracy of a wrist-worn and ear-worn activity monitor. Nineteen healthy adults completed 5-minute bouts of running at 5.0 mph and 6.0 mph on a treadmill, while wearing the Fitbit Charge on the wrist and Jabra Sport Pulse Wireless Earbuds (GN Netcom, Ballerup, Denmark) in the ears. The Fitbit Charge and Jabra earbuds significantly underestimated EE by 14% and 16%, respectively, compared to indirect calorimetry during the 5.0 mph running speed ($P<0.05$). Both the Fitbit Charge and Jabra earbuds displayed a greater underestimation in EE at the 6.0 mph running speed (Fitbit: -23%, Jabra: -23%) compared to the 5.0 mph running speed (Fitbit: -14%, Jabra: -16%). The Jabra earbuds displayed greater variability than the Fitbit Charge at both running speeds. The Jabra earbuds were not

more accurate than the wrist-worn Fitbit Charge at measuring EE and both devices displayed greater error in EE measurements at faster running speeds (35).

Diaz et al. (37) examined the accuracy of a wrist-worn and hip-worn activity monitor for running. Twenty-three healthy adults ran at a speed of 5.2 mph on a treadmill for 6 minutes, while wearing the Fitbit Flex at the wrist and the Fitbit One at the hip. During the run, all devices overestimated EE compared to indirect calorimetry. Between the two devices, the Fitbit One (hip-worn) displayed greater accuracy than the Fitbit Flex (wrist-worn) at measuring EE (37).

In a follow-up study, Diaz et al. (36) had subjects wear the Fitbit Flex on the wrist and Fitbit One at the hip, as well as an additional Fitbit One on the upper torso. Thirteen healthy females ran at 5.2 mph on a treadmill for 6 minutes, while wearing the three devices. The Fitbit One at the hip (-3%) displayed greater accuracy in EE measurements for running than the Fitbit One at the upper torso (11%) and the Fitbit Flex (25%) compared to indirect calorimetry. Diaz et al. demonstrated that although the Fitbit One worn at the upper torso was less accurate at measuring EE than wearing it at the hip, it was still more accurate than measuring EE at the wrist by the Fitbit Flex for running (36).

Studies examining the accuracy of consumer-grade activity monitors for running at 5.0-8.5 mph have determined most activity monitors to be inaccurate at measuring EE (27,28,34,35,38,39,43–48,50,51), with some activity monitors displaying a %RE of >10% (35,43) and a MAPE of >10% (34,38,43,50,51). Activity monitors have been demonstrated to display greater inaccuracy at faster running speeds, including up to maximal-intensity exercise (35,48,51,52). In addition, Diaz et al. demonstrated in two

separate studies that a hip-worn activity monitor was more accurate than a wrist-worn activity monitor at measuring EE for running, while Dondzila and Garner did not observe greater accuracy in EE measurements by an ear-worn activity monitor compared to a wrist-worn monitor (35–37).

2.1.3. Elliptical Exercise

Three studies have incorporated elliptical exercise into their physical activity protocol for the assessment of EE accuracy in consumer-grade activity monitors (30,32,45). However, two of the three studies only examined accuracy in EE during all activities combined and did not report accuracy of the devices during individual types of activity (30,32). As a result, the accuracy of these devices at measuring EE during the different activities performed, such as elliptical exercise, is unknown.

However, Stackpool et al. (45) assessed and reported the accuracy of five different activity monitors at measuring EE during 20 minutes of self-paced exercise on an elliptical cross-trainer. Twenty healthy adults wore the Adidas miCoach (Adidas AG, Herzogenaurach, Germany), BodyMedia FIT CORE (BodyMedia, Pennsylvania, USA), Fitbit Ultra, Jawbone UP, and Nike FuelBand (Nike, Inc., Oregon, USA) while performing elliptical exercise. Devices were worn according to manufacturers' instructions, but the exact wear location for each device was not reported. The Adidas miCoach was unable to detect elliptical exercise, and therefore, no data was reported for this device. The BodyMedia FIT CORE and the Nike FuelBand significantly underestimated EE by 20% and 27%, respectively, compared to indirect calorimetry ($P<0.05$). A weak relationship was observed between the devices and indirect

calorimetry, with Pearson correlation coefficients of 0.47, 0.41, 0.40, and 0.08 displayed for the BodyMedia FIT CORE, Fitbit Ultra, Jawbone UP, and Nike FuelBand, respectively. The NikeFuel Band displayed the greatest underestimation and the smallest correlation with indirect calorimetry out of all of the devices assessed (45).

2.1.4. Cycling

Eleven studies have assessed the accuracy of consumer-grade activity monitors for measuring EE during cycling (27,28,30–32,38,43,46,47,50,53). Seven of these studies have examined and reported the accuracy of devices during different types of activities, while the remaining four only examined accuracy in EE during all activities performed combined and did not report accuracy during individual activities, such as cycling (28,30–32).

Nelson et al. (50) had 16 healthy adults cycle for 5 minutes at a self-selected intensity on a cycle ergometer, while wearing the Fitbit Flex and Jawbone UP24 (Jawbone, California, USA) on the wrist and the Fitbit One and Fitbit Zip (Fitbit, Inc., California, USA) on the hip. All devices significantly underestimated EE during cycling by 37-59% ($P=0.025 - <0.001$). All devices displayed a high MAPE during cycling compared to indirect calorimetry. The MAPE in EE measurements during cycling were 43%, 44%, 46%, and 57% for the Fitbit One, Fitbit Flex, Fitbit Zip, and Jawbone UP24, respectively (50).

Sasaki et al. (47) examined the accuracy of the Fitbit at measuring EE during 6 minutes of steady-state cycling at 50 W on a cycle ergometer in 20 healthy adults. The Fitbit significantly underestimated EE by 104% compared to indirect calorimetry

($P < 0.05$), while a %RE of approximately $\leq 40\%$ was demonstrated by the Fitbit during other activities, such as activities of daily living (ADL), ambulatory activities, and sports (47).

Chowdhury et al. (38) had 30 healthy adults perform 10 minutes of steady-state cycling on a cycle ergometer, with females cycling at 75 W and males cycling at 100 W. All subjects wore the Apple Watch (Apple Inc., California, USA), Fitbit Charge HR, Jawbone UP24, and Microsoft Band (Microsoft Corporation, Washington, USA) on the wrists while cycling. All devices displayed high MAPE compared to indirect calorimetry (20-82%). The MAPE in EE measurements during cycling was 20%, 38%, 53%, and 82% for the Apple Watch, Microsoft Band, Fitbit Charge HR, and Jawbone UP24, respectively (38).

Instead of having subjects cycle at a single intensity, Shcherbina et al. (46) had subjects complete both low- and high-intensity cycling. Sixty healthy adults performed 5 minutes of low-intensity (50-100 W) cycling and 5 minutes of high-intensity (80-225 W) cycling on a cycle ergometer, while wearing five wrist-worn activity monitors: Apple Watch, Basis Peak (Intel Corporation, California, USA), Fitbit Surge (Fitbit, Inc., California, USA), Microsoft Band, and PulseOn (PulseOn, Espoo, Finland). EE data from both cycling intensities was pooled together. The median absolute percentage error ranged from ~25-100% on all devices. The Fitbit Surge displayed the lowest median absolute percentage error (~25%), while the PulseOn displayed the highest (~100%) (46).

Pribyslavskaya et al. (43) assessed the accuracy of EE measurements during moderate- and vigorous-intensity cycling. Thirty-four healthy adults completed 10-

minute bouts of cycling at moderate-intensity (40-59% of heart rate reserve [HRR]) and vigorous-intensity (60-84% of HRR) on a cycle ergometer, while wearing the Fitbit Surge and Garmin vívofit on the wrists. Both activity monitors underestimated EE for cycling at moderate-intensity (Fitbit: 20%, Garmin: 6%) and vigorous-intensity (Fitbit: 23%, Garmin: 0%). The Fitbit Surge displayed a MAPE of 28% for both moderate- and vigorous-intensity cycling, while the Garmin vívofit displayed a MAPE of 19% for moderate-intensity cycling and 16% for vigorous-intensity cycling (43).

Woodman et al. (27) conducted the only study to assess the accuracy of activity monitors during both indoor and outdoor cycling. Twenty-five healthy adults completed 5 minutes of indoor cycling on a cycle ergometer at 100 W and 26 adults completed 5 minutes of cycling outdoors at a self-selected speed, while wearing the Basis Peak and Garmin vívofit on the wrists and three Withings Pulse (Withings, Issy-les-Moulineaux, France) activity monitors (one on shirt collar, one on hip, and one on wrist). All devices significantly underestimated EE during both indoor and outdoor cycling compared to indirect calorimetry ($P < 0.05$). The Basis Peak displayed the least underestimation for indoor (7.0 ± 5.1 kcal/min) and outdoor cycling (5.4 ± 3.5 kcal/min), while the Withings Pulse on the wrist displayed the greatest underestimation for both indoor (1.2 ± 2.0 kcal/min) and outdoor cycling (1.2 ± 2.0 kcal/min) compared to indirect calorimetry (Indoor: 9.2 ± 1.2 kcal/min, Outdoor: 8.7 ± 2.6 kcal/min). There was no significant difference in EE measurements from the Withings Pulse devices placed at the collar, hip, or wrist during indoor cycling ($P < 0.05$). However, EE measurements from the Withings Pulse at the shirt collar and the hip were both significantly higher than the device at the wrist ($P < 0.001$), but not significantly different from each other ($P > 0.05$).

during outdoor cycling, indicating greater accuracy by the Withings Pulse at measuring EE at the shirt collar and hip than the wrist (27).

While most studies have examined the accuracy of activity monitors during steady-state cycling, Boudreaux et al. (53) completed a study assessing EE accuracy during a graded exercise test. Fifty healthy adults performed a graded exercise test on a cycle ergometer, while wearing seven different activity monitors: Apple Watch 2 (Apple Inc., California, USA), Fitbit Blaze (Fitbit, Inc., California, USA), Fitbit Charge 2 (Fitbit, Inc., California, USA), Garmin vivosmart HR (Garmin International, Inc., Kansas, USA), Polar A360 (Polar Electro, Kempele, Finland), Polar H7 (Polar Electro, Kempele, Finland), and TomTom Touch (TomTom International, Amsterdam, Netherlands). Subjects wore six devices on the wrists (three per wrist) and the Polar H7 on the chest. They began cycling at 50 rpm and 50 W and every 2 minutes the power output was increased by 25 W until exhaustion. EE measurements from all devices, except the Garmin vivosmart HR, were significantly different from indirect calorimetry ($P<0.05$). All the devices tested displayed weak ICC with the indirect calorimeter ($r=0.18-0.41$). All devices also displayed high MAPE compared to indirect calorimetry (21-75%). The Apple Watch 2 displayed the lowest MAPE (21%), while the Fitbit Charge 2 displayed the highest MAPE (75%). The Fitbit Blaze and Fitbit Charge 2 underestimated EE in 84% and 80% of subjects, respectively, while the remaining five devices overestimated EE in 52-98% of subjects during cycling (53).

Twenty-one different consumer-grade activity monitors have been tested for accuracy in EE during cycling. All devices tested by Nelson et al., Chowdhury et al., Pribyslavska et al., and Boudreaux et al. displayed high MAPE values, ranging from 16-

82% (38,43,50,53). While Sasaki et al. and Shcherbina et al. did not report MAPE values for the activity monitors tested, Sasaki et al. reported a mean underestimation in EE by 104% for the Fitbit and Shcherbina et al. reported median absolute percentage errors ranging from ~25-100% for the devices tested (46,47). Woodman et al. observed a significant underestimation in EE during both indoor and outdoor cycling by three different activity monitors and greater accuracy in EE measured at the shirt collar and hip compared to the wrist during outdoor cycling (27). In addition, two studies assessed the accuracy of the Jawbone UP24 during cycling, and in both studies, this device displayed a higher MAPE (57-82%) than any of the other activity monitors tested in each study (38,50).

2.1.5. Resistance Exercise

Although the Canadian Society for Exercise Physiology, American College of Sports Medicine, Diabetes Canada, and ADA have all recommended performing resistance exercise on at least 2 days of the week, only two studies have examined the accuracy of consumer-grade activity monitors during resistance exercise (15,16,40,53–55).

Boudreaux et al. (53) had subjects perform 3 circuits of 10 repetitions of leg curls, chest press, leg extensions, and latissimus dorsi pulldowns, while wearing seven different activity monitors. Fifty healthy adults wore the Apple Watch 2, Fitbit Blaze, Fitbit Charge 2, Garmin vivosmart HR, Polar A360, and TomTom Touch on the wrists and the Polar H7 on the chest. All of the devices tested displayed weak ICC with the indirect calorimeter ($r=0.02-0.18$). All devices displayed high MAPE compared to indirect

calorimetry (43-57%). The Apple Watch displayed the lowest MAPE (43%), while the Garmin vivosmart HR displayed the highest MAPE (57%). All devices overestimated EE in 60-90% of subjects during resistance exercise (53).

On the other hand, Bai et al. (40) conducted a study where they aimed to simulate real-world conditions by allowing subjects to perform 25 minutes of resistance exercise using self-selected exercises, weights, sets, and repetitions. Subjects performed any combination of 12 different exercises: rotary torso, abdominal crunch, lower back, vertical traction, shoulder press, arm curl, arm extension, leg press, leg curl, leg extension, multi-hip, and chest press. Fifty-two healthy adults performed all exercises while wearing the Fitbit Flex, Jawbone UP24, Misfit Shine (Misfit, California, USA), and Nike+ FuelBand SE (Nike, Inc., Oregon, USA) on the wrists. The Fitbit Flex (25.5 kcal) and Nike+ FuelBand SE (26.3 kcal) displayed the smallest mean bias compared to indirect calorimetry, followed by the Misfit Shine (30.8 kcal) and Jawbone UP24 (47.2 kcal). All devices underestimated EE during resistance exercise. The lowest MAPE was displayed by the Fitbit Flex (32%) and the highest MAPE was displayed by the Jawbone UP24 (53%) (40).

In total, 11 different consumer-grade activity monitors were assessed for accuracy in EE during resistance exercise between both studies, and all devices tested displayed high MAPE, ranging from 32-57%. However, Bai et al. reported that all devices tested underestimated EE, while Boudreaux et al. observed an overestimation in EE in 60-90% of subjects during resistance exercise (40,53).

2.1.6. Sports and Sports-Related Exercise

Four studies have examined the accuracy of consumer-grade activity monitors during sports or sports-related exercise (30,45,47). Lee et al. (30) had subjects play basketball and tennis during a 69-minute exercise protocol, but did not report the accuracy of devices during individual activities. As a result, the accuracy of the devices tested in this study during basketball and tennis are unknown.

However, Sasaki et al. (47) examined and reported the accuracy of the Fitbit while subjects performed 6 minutes of basketball, golf, and tennis, for a total time of 18 minutes. The Fitbit was worn at the hip by 20 healthy adults while performing the different sports. The Fitbit displayed an overall underestimation in EE during sports, with a mean bias of -2.1 kcal/6 min (95% LoA: -26.0 to 22.0 kcal/6 min) compared to indirect calorimetry. However, for individual sports, the Fitbit underestimated EE during tennis by ~15%, while overestimating EE during basketball by ~20%, and displaying almost no mean difference (~0%) in EE during golf (47).

Wahl et al. (39) had 20 healthy adults perform 5 minutes of intermittent running simulating a soccer game, while wearing various activity monitors: Beurer AS80, Fitbit Charge, Fitbit Charge HR, Garmin Forerunner 920XT, Garmin vívoactive, Garmin vívofit, Garmin vívosmart, Polar Loop, and Withings Pulse O_x. One of each device was worn on the wrist, while an additional Withings Pulse O_x monitor was worn at the hip. The Garmin vívosmart, Fitbit Charge, Fitbit Charge HR, and Polar Loop all overestimated EE by 1-49% during the intermittent running, while the Beurer AS80, Garmin Forerunner 920XT, Garmin vívoactive, Garmin vívofit, and Withings Pulse O_x at the hip and wrist all underestimated EE by 2-26% compared to indirect calorimetry. The

Garmin vívoactive (-1%), Fitbit Charge (2%), Garmin vivosmart (2%), Polar Loop (6%), and Garmin Forerunner 920XT (-9%) were the only devices to display a %RE of $\leq 10\%$. The Garmin vívofit ($r=0.54$), Fitbit Charge ($r=0.58$), and Garmin vívoactive ($r=0.74$) all displayed a moderate ICC with indirect calorimetry. All other devices displayed poor ICC ($r=0.00-0.43$) with indirect calorimetry for intermittent running (39).

Meanwhile, Stackpool et al. (45) had subjects perform 20 minutes of basketball-related activities, including agility ladder drills, basketball free throws, T drills, and a basketball half-court lay-up drill, while wearing the Adidas miCoach, BodyMedia FIT CORE, Fitbit Ultra, Jawbone UP, and Nike FuelBand. The devices were worn by 20 healthy adults and were worn according to manufacturers' instructions, but the exact wear location for each device was not reported. All of the devices tested significantly underestimated EE by 14-60% compared to indirect calorimetry ($P<0.05$). The %RE was -14%, -17%, -18%, -30%, and -60% for the Nike FuelBand, Fitbit Ultra, BodyMedia FIT CORE, Jawbone UP, and Adidas miCoach, respectively. The Fitbit Ultra ($r=0.67$), Jawbone UP ($r=0.65$), Adidas miCoach ($r=0.57$), and BodyMedia FIT CORE ($r=0.56$) all displayed moderate relationships with indirect calorimetry, while the Nike FuelBand ($r=0.47$) displayed a weak relationship (45).

Of the 15 consumer-grade activity monitors tested, 11 devices displayed a mean underestimation in EE during sports and sports-related exercise (39,45,47). However, while Sasaki et al. examined an overestimation in EE during basketball by the Fitbit, Stackpool et al. observed an underestimation in EE during basketball-related activities by five different activity monitors (45,47).

2.2. ACCURACY OF ACTIVITY MONITORS DURING DAILY LIVING AND SEDENTARY ACTIVITIES

2.2.1. Activities of Daily Living

Seven studies have assessed the accuracy of consumer-grade activity monitors at measuring EE during activities of daily living (ADL) (27,31,32,38,41,47,50). However, only five of these studies have examined and reported the accuracy of devices during individual types of activities, including ADL.

Nelson et al. (50) had subjects perform eight ADL (standing, dusting, sweeping, vacuuming, folding laundry, making bed, picking up items from floor, and gardening) for 5 minutes each, while wearing the Fitbit Flex, Fitbit One, Fitbit Zip, and Jawbone UP24. Fifteen healthy adults performed the activities, while wearing the Fitbit Flex and Jawbone UP24 on the wrists and the Fitbit One and Fitbit Zip on the hip. All of the devices significantly underestimated EE during ADL by 27-34% ($P<0.001$), except the Fitbit Flex, which was not significantly different from indirect calorimetry. All devices displayed high MAPE during ADL. The MAPE in EE measurements during ADL were 21%, 27%, 35%, and 36% for the Fitbit Flex, Fitbit One, Jawbone UP24, and Fitbit Zip respectively (50).

Chowdhury et al. (38) had 30 healthy adults complete 15 minutes of ADL, including loading and unloading a dishwasher, sweeping, and self-paced ascending and descending of stairs, while wearing the Apple Watch, Fitbit Charge HR, Jawbone UP24, and Microsoft Band at the wrists. MAPE in EE ranged from 15-62% across all devices during all three activities. The Fitbit Charge HR displayed the lowest MAPE during the loading and unloading of a dishwasher (23%) and walking up and down stairs (15%),

while the Apple Watch displayed the lowest MAPE during sweeping (30%). The Microsoft Band displayed the highest MAPE in EE during all three activities, with a MAPE of 54%, 45%, and 62% during the loading and unloading of a dishwasher, sweeping, and walking up and down stairs, respectively (38).

Bai et al. (41) had 39 healthy adults perform 25 minutes of ADL, including folding laundry, sweeping, moving light boxes, stretching, and slow walking, while wearing the Apple Watch and Fitbit Charge HR on the wrist. The Apple Watch displayed a small overestimation in EE by 4%, while the Fitbit Charge HR displayed a large underestimation in EE by 60% compared to indirect calorimetry. Although Chowdhury et al. (38) observed a MAPE of 29-34% by the Apple Watch, Bai et al. observed a lower overall MAPE of 23% by this device during ADL. On the other hand, Chowdhury et al. reported that the Fitbit Charge HR displayed the lowest MAPE (15-23%) during two out of three ADL performed compared to three other activity monitors, while Bai et al. observed a high overall MAPE of 61% by the Fitbit Charge HR during ADL. The MAPE observed by Bai et al. for the Fitbit Charge HR (61%) was over two times greater than the MAPE displayed by the Apple Watch (23%). In addition, the Fitbit Charge HR ($r=0.41$, $P<0.05$) displayed a weaker correlation with the indirect calorimeter than the Apple Watch ($r=0.66$, $P<0.01$) (41).

Woodman et al. (27) had subjects perform 15 minutes of ADL. Twenty-five healthy adults folded clothes, swept the floor, and ascended and descended stairs, while wearing the Basis Peak and Garmin vívofit on the wrists and three Withings Pulse activity monitors (one on shirt collar, one on hip, and one on wrist). All devices significantly underestimated EE during all ADL compared to indirect calorimetry

($P < 0.05$), except for the Basis Peak, which overestimated EE for walking up and down stairs. All devices reported mean EE as 2.0-2.8 kcal/min for folding clothes and sweeping, while the indirect calorimeter reported mean kcal as 3.1 and 4.2 kcal/min for folding clothes and sweeping, respectively. For ascending and descending stairs, the indirect calorimeter reported mean EE as 9.1 kcal/min, while the EE measurements from the devices ranged from 3.9 kcal/min by the Withings Pulse on the wrist to 10.3 kcal/min by the Basis Peak. During all ADL, EE measurements from the Withings Pulse at the shirt collar and the hip were both significantly different than the device at the wrist ($P < 0.001$), but not significantly different from each other ($P > 0.05$) (27).

Sasaki et al. (47) had 20 healthy adults perform 30 minutes of ADL, including dusting, gardening, laundry, raking, and vacuuming for 6 minutes each, while wearing the Fitbit at the hip. The Fitbit displayed an overall underestimation in EE during ADL, with a mean bias of -3.1 kcal/6 min (95% LoA: -11.0 to 5.2 kcal/6 min) compared to indirect calorimetry. Underestimation in EE occurred during all activities, except for dusting, which displayed an overestimation in EE measurements. Bland-Altman plots revealed narrower 95% LoA for ADL (difference: 16.2 kcal) compared to ambulatory activities (difference: 47.0 kcal) and sports (difference: 48.0 kcal) (47).

Across these five studies, 11 different activity monitors were assessed for accuracy in EE during ADL. Seven out of 10 devices underestimated EE during ADL (27,38,41,47,50). Bai et al. (41) observed a lower MAPE in EE measurements by the Apple Watch and a higher MAPE by the Fitbit Charge HR compared to Chowdhury et al. (38). Woodman et al. (27) examined a significant difference in EE reported by devices worn at the shirt collar and hip compared to the wrist, but all devices were still

significantly different from indirect calorimetry, regardless of device wear location. In addition, Sasaki et al. (47) observed narrower 95% LoA for the Fitbit during ADL compared to ambulatory activities and sports, indicating greater accuracy in EE measurements by the Fitbit during ADL compared to the other activities.

2.2.2. Sedentary Activities

In addition to assessing the accuracy of different consumer-grade activity monitors during exercise and ADL, five studies have also examined accuracy during various sedentary activities. Two studies only examined accuracy in EE during all activities performed combined and did not report accuracy of the devices tested during individual types of activity. Consequently, the accuracy of these devices at measuring EE during different activities, such as sedentary activities, is unknown.

Nelson et al. (50) had 10-30 healthy adults perform different sedentary activities, including lying down (n=30), computer use (n=15), watching television (n=15), writing (n=10), playing cards (n=10), and reading (n=10), while wearing the Fitbit Flex and Jawbone UP24 on the wrist and the Fitbit One and Fitbit Zip on the hip. All activities were performed for 5 minutes each, with the exception of lying, which was performed for 10 minutes. The Jawbone UP24 significantly underestimated EE by 8% ($P=0.013$) compared to indirect calorimetry. The MAPE in EE measurements during sedentary activities were 13%, 14%, 16%, and 17% for the Fitbit One, Fitbit Flex, Fitbit Zip, and Jawbone UP24, respectively. All activity monitors had the lowest MAPE during sedentary activities (13-17%) compared to ADL (21-36%) and ambulatory activities (16-36%) (50).

Bai et al. (40) assessed the accuracy of EE measurements during 20 minutes of sedentary activities, including reading, using a cell phone, typing, watching videos, and listening to music/radio. Fifty-two healthy adults completed the activities while wearing the Fitbit Flex, Jawbone UP24, Misfit Shine, and Nike+ FuelBand SE on the wrists. The Fitbit Flex (5.5 kcal) and Misfit Shine (5.7 kcal) displayed the smallest mean bias, followed by the Nike+ FuelBand SE (6.5 kcal) and Jawbone UP24 (11.0 kcal) compared to indirect calorimetry. All devices underestimated EE during sedentary activities. MAPE was lowest for the Misfit Shine (18%) and Nike+ FuelBand SE (20%) during sedentary activities compared to indirect calorimetry (40).

Bai et al. (41) completed a follow-up study where subjects also completed 20 minutes of sedentary activities, such as working on a computer, reading, and using a cell phone. In this study, they assessed the accuracy of the Apple Watch and Fitbit Charge HR at measuring EE at the wrist in 39 healthy adults. As opposed to their previous study, Bai et al. observed an overestimation in EE by both devices during sedentary activities compared to indirect calorimetry in this study, with the Apple Watch overestimating by 9% and the Fitbit Charge HR overestimating by 21%. Both devices displayed a high MAPE (Apple: 22%, Fitbit: 35%) and a weak relationship with indirect calorimetry, with Pearson correlation coefficients of 0.45 ($P<0.01$) and 0.40 ($P<0.05$) for the Apple Watch and Fitbit Charge HR, respectively (41).

In total, eight consumer-grade activity monitors were assessed for accuracy in EE during sedentary activities across three different studies, with five of the eight devices displaying an underestimation in EE (40,41,50). In addition, Nelson et al.

observed the lowest MAPE in EE measurements during sedentary activities compared to ADL and ambulatory activities (50).

2.3. ACCURACY OF ACTIVITY MONITORS IN INDIVIDUALS WITH DIABETES

Although no studies to date have assessed the accuracy of consumer-grade activity monitors exclusively in individuals with T1D, one study examined the accuracy of consumer-grade and research-grade activity monitors in a combination of individuals with and without T1D, while another study assessed the accuracy of research-grade activity monitors in individuals with T1D and type 2 diabetes (T2D).

Yavelberg et al. (56) had 8 individuals with T1D and 17 individuals without T1D perform 40 minutes of steady-state exercise on a treadmill at $\leq 50\%$ of their maximum volume of oxygen consumption (VO_2 max) and a 40-minute exercise circuit ranging from 50-100% of VO_2 max. The exercise circuit involved performing variations of high knees, squats with a medicine ball, jumping jacks, push-ups, and forearm planks. All subjects wore the Metria IH1 arm patch (Avery Dennison Corporation, California, USA) (research-grade) and a subset of subjects also wore the Garmin vívofit 2 (n=5-7) (consumer-grade) and Mio Fuse (PAI Health Inc., British Columbia, Canada) (n=5-6) (consumer-grade) during exercise. The Metria IH1 displayed a mean bias of 0.3 kcal/min (95% LoA: -2.9 to 3.5 kcal/min) during the steady-state exercise compared to indirect calorimetry. During the moderate-to-vigorous intensity (50-75% VO_2 max) portion of the circuit, the Metria IH1 displayed a mean bias of 0.6 kcal/min (95% LoA: -2.6 to 3.9 kcal/min), while displaying a mean bias of -1.8 kcal/min (95% LoA: -7.2 to 3.7 kcal/min) during the vigorous-to-maximum intensity (75-100% VO_2 max) portion of the

circuit. The Metria IH1 displayed a 4% and 6% overestimation in EE during the steady-state exercise and moderate-to-vigorous intensity portion of the exercise circuit, respectively, and displayed a 13% underestimation in EE during the vigorous-to-maximum intensity portion of the circuit ($P<0.001$). The Garmin vívofit 2 underestimated EE during both the steady-state exercise ($P<0.001$) and the exercise circuit ($P>0.05$) compared to indirect calorimetry, while the Mio Fuse overestimated during both types of exercise ($P>0.05$) (56).

Machac et al. (57) assessed the accuracy of two research-grade activity monitors in individuals with T1D and T2D while walking on a treadmill. Nineteen adults with T1D or T2D walked for 15-minute bouts at 1.9, 2.5, and 3.1 mph and 0%, 0%, and 5% incline, respectively, while wearing the OMRON HJ-720 (OMRON Corporation, Kyoto, Japan) and the SenseWear Armband Pro3 (BodyMedia, Pennsylvania, USA). The OMRON HJ-720 and SenseWear Armband Pro3 both displayed an overestimation in EE by 71% and 81% during level walking at 1.9 mph and by 76% and 78% for level walking at 2.5 mph, respectively. During higher intensity walking at 3.1 mph and 5% incline, the OMRON HJ-720 and SenseWear Armband Pro3 both displayed an underestimation in EE by 7% and 8%, respectively (57).

The studies conducted by Yavelberg et al. (56) and Machac et al. (57) are the only studies to have assessed the accuracy in EE by activity monitors in individuals with T1D. Yavelberg et al. observed increased error in EE measurements at higher exercise intensities by the Metria IH1, which is consistent with findings from studies performed with consumer-grade activity monitors in healthy adults (51,52). However, Machac et al.

displayed greater accuracy in EE at a higher walking intensity, which is contradictory to the findings from studies conducted in healthy adults (28,33,48,49).

2.4. SUMMARY

Overall, studies examining the accuracy of consumer-grade activity monitors found most devices to be inaccurate at measuring EE during various types of exercise, ADL, and sedentary activities (27–29,33–35,38,39,42–48,50,51,53). In most of these studies, activity monitors displayed a %RE of >10% (35–37,42,45,47) and a MAPE of >10% (34,38,40,41,43,47,50,51,53). Many studies also discovered that as exercise intensity increased, accuracy of EE measurements by activity monitors decreased (28,33,35,48,49,51,52).

Although a variety of studies have been conducted to assess the accuracy of consumer-grade activity monitors, more research in this field is still needed. Many different activity monitors have already been tested, but new models of activity monitors are being released to the consumer market every year. These new devices also need to be assessed for accuracy in EE measurements and other physiological metrics. The majority of the existing studies conducted in this field focus on the accuracy of activity monitors during steady-state walking and running, while only a few studies have examined accuracy during other forms of physical activity. More studies need to be conducted to assess the accuracy of activity monitors during common forms of physical activity, such as elliptical exercise, cycling, resistance exercise, and sports. In addition, almost all of the studies conducted to date have examined the accuracy of activity monitors in healthy adults, with only two studies including individuals with T1D as

subjects. However, even within those two studies, data from individuals with and without T1D were pooled together by Yavelberg et al. (56), while Machac et al. (57) only assessed the accuracy of research-grade activity monitors in individuals with T1D and T2D, but did not assess any consumer-grade monitors. Research needs to be conducted to assess the accuracy of consumer-grade activity monitors exclusively in individuals with T1D, as well as in other non-healthy populations.

3. STUDY OVERVIEW, OBJECTIVES & HYPOTHESES

3.1. STUDY OVERVIEW

The purpose of this project was to determine the accuracy and utility of various consumer-grade, wrist-worn activity monitors for measuring physical activity metrics in adults with and without T1D. Two similar, but separate, studies were conducted for this project. Study A was a pre-pilot study that was conducted in 20 healthy adults using two common consumer-grade, wrist-worn activity monitors (Fitbit Charge 2 [FC2], Garmin vívosmart HR+ [GVHR+]) during 4 types of common physical activity. Study B was a follow-up study that was conducted in 15 adults (9 T1D, 6 healthy) using two newer consumer-grade, wrist-worn activity monitors (Fitbit Ionic [FI], Garmin vívosmart 3 [GV3]) during 5 types of common physical activity.

Study A and B were the first phases of a larger, multicentre, long-term project, known as the Type 1 Diabetes and Exercise Initiative (T1DEXI), supported by the Helmsley Trust. One of the long-term goals of this research are to find activity monitors that are accurate enough at measuring EE and other physical activity parameters to be incorporated into an AP system.

3.2. PRIMARY OBJECTIVES

The primary objective of Study A was to determine if the FC2 and GVHR+ were accurate (i.e. $\leq 10\%$ error of a laboratory standard) (50,51,53) at measuring EE during different types of physical activity in healthy adults.

The primary objective of Study B was to determine if the FI and GV3 were accurate (i.e. $\leq 10\%$ error of a laboratory standard) (50,51,53) at measuring EE during different types of physical activity in adults with and without T1D.

3.3. SECONDARY OBJECTIVES

The secondary objectives of Study A were:

- to examine the %RE in metabolic equivalents (METs) for the FC2 and GVHR+
- to examine the %RE in volume of oxygen consumption (VO_2) for the FC2 and GVHR+

The secondary objectives of Study B were:

- to examine glucose changes over the time of exercise
- to examine the time spent in euglycemia (3.9-10.0 mmol/L) during and for 12 hours post-exercise
- to examine the time spent in hypoglycemia (< 3.9 mmol/L) during and for 12 hours post-exercise
- to examine the time spent in hyperglycemia (> 10.0 mmol/L) during and for 12 hours post-exercise

3.4. HYPOTHESES

For Study A, it was hypothesized that the FC2 and GVHR+ would display a similar level of accuracy to each other, but would not achieve an acceptable level (i.e.

$\leq 10\%$ %RE) of accuracy at measuring EE during different types of physical activity in healthy adults (35–37,42,45,47,53).

For Study B, it was hypothesized that the newer, consumer-grade activity monitors, the FI and GV3, would display greater accuracy than the FC2 and GVHR+, but would not achieve an acceptable level (i.e. $\leq 10\%$ %RE) of accuracy at measuring EE during different types of physical activity in adults with and without T1D (35–37,42,45,47,53). Although this has not previously been assessed, we hypothesized that there would no difference in the accuracy of EE measurements between individuals with and without T1D. In addition, we hypothesized that there would be a greater decrease in BG in individuals with T1D during exercise visit 1 (three moderate- to vigorous-intensity activities) compared to visit 2 (one low- and one high-intensity activity) (58–63).

4. STUDY A: PRE-PILOT STUDY

4.1. METHODS

4.1.1. Participants

Twenty healthy adults (10 subjects at York University, 10 subjects at Oregon Health and Science University) were recruited for this study. All participants provided written informed consent before taking part in the study. All participants were screened for cardiovascular complications using the Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) (64). Physical activity levels were assessed in all participants using the International Physical Activity Questionnaire (IPAQ) (65). This study was approved by the Human Participants Review Sub-Committee at York University (Toronto, Ontario) (Certificate #: e2018-010) and the Institutional Review Board at Oregon Health and Science University (Portland, Oregon).

4.1.2. Study Protocol

Participants attended the research laboratory on two separate occasions, separated by a minimum of 24 hours. Anthropometrics were gathered on all participants at the beginning of their first laboratory visit. After completing this, participants performed a progressive aerobic exercise test to peak oxygen consumption on a treadmill (VO₂ peak-T) or cycle ergometer (VO₂ peak-C) and a resistance exercise circuit (described in more detail in section 4.1.2.1. *Exercise Visit #1*). During the second laboratory visit, participants performed a series of common activities of daily living (ADL) and high-intensity interval training on a treadmill (HIIT-T) or cycle ergometer (HIIT-C) (described in more detail in section 4.1.2.2. *Exercise Visit #2*). All participants wore a

portable metabolic system (Cosmed K4b²/K5), a HR monitor (Polar A300 and Polar H7), and two wrist-worn, consumer-grade activity monitors (FC2, GVHR+) during all exercises (described in more detail in section 4.1.3. *Devices*).

During the first laboratory visit and prior to starting exercise, anthropometrics were gathered on all participants. Height was measured using a wall-mounted stadiometer to the nearest 0.1 cm. Participants were instructed to remove their shoes and stand up straight against the stadiometer with their back and heels touching the wall. They were instructed to take a deep breath and exhale, after which point an investigator lowered the headpiece to the top of the participant's head. The participant was asked to step away from the wall and their height was recorded. Weight and body fat percentage (BF%) were measured using the Tanita BF-350 Total Body Composition Analyzer scale (Tanita Corporation, Tokyo, Japan) to the nearest 0.1 kg and 0.1%, respectively. Participant's sex, age, and height were entered into the Tanita scale. Once this information had been entered into the scale, participants were instructed to remove their shoes and socks, step on the scale, and remain standing on the scale until their weight and BF% were displayed on the screen and recorded. Waist circumference (WC) and wrist circumference were measured using a soft vinyl tape measure by one of the study investigators to the nearest 0.1 cm. WC was measured horizontally around the abdomen at the level of the iliac crest (66). Wrist circumference was measured adjacent and proximal to the styloid bone of the ulna on the dominant hand. Blood pressure and resting HR were measured using the OMRON M7 Intelli IT blood pressure monitor and OMRON Intelli Wrap HEM-FL31 blood pressure cuff (OMRON Corporation, Kyoto, Japan). Participants were seated in a chair with the blood pressure cuff wrapped around

their left arm, which was placed on a table at chest height next to them. Participants were instructed to relax, not speak, and keep their legs uncrossed and feet flat on the floor during the measurements (67,68). Duplicate blood pressure and resting HR measurements were taken 60 seconds apart and then averaged together.

4.1.2.1. Exercise Visit #1

Participants arrived at the lab in exercise-appropriate clothing (e.g. t-shirt, tank top, shorts, leggings, etc.). After initial anthropometrics had been measured and recorded, two consumer-grade activity monitors (FC2, GVHR+) were placed on opposing wrists (as per manufacturer's instructions) on the participant in a randomized and counterbalanced method. In addition, all participants wore a portable metabolic system (Cosmed K4b²/K5) and a HR monitor (Polar H7 and A300) during all exercises.

All participants were given a 5-minute rest period for baseline assessment and recovery, before and after each exercise, respectively. Participants were instructed to remain seated during these assessments. Once the Cosmed K4b²/K5 and both activity monitors had been placed into the appropriate activity mode (Appendix A), study investigators started recording on all devices, which began the 5-minute, pre-exercise rest. Five minutes after each exercise ended, the devices were stopped and the 5-minute, post-exercise rest ended.

The first exercise participants performed was the VO₂ peak-T/VO₂ peak-C. Participants were assigned to a mode of exercise (i.e. treadmill or cycle ergometer) in a randomized and counterbalanced method by investigators. For the VO₂ peak-T, each participant began with a 4-minute walking warm-up (3.0 mph, 0% grade for 2 min and

then 3.0 mph, 5% grade for 2 min) on the Spirit CT850 treadmill (Dyaco International, Taipei, Taiwan). After the warm-up, participants self-selected a comfortable running speed between 4 to 6 mph, and subsequently, the treadmill incline was increased by 2% every 2 minutes until the participants reached volitional exhaustion (52). After reaching exhaustion, participants completed a 2-minute walking cool down at 2.5 mph and 0% incline. For the VO_2 peak-C, each participant began with a 2-minute warm-up (60 rpm, 0 W) on the Lode Corival cycle ergometer (Lode BV, Groningen, Netherlands). After the warm-up, the cadence was maintained at 60 rpm and the power output was increased every 2 minutes by 25 W for females and 30 W for males until the participants reached volitional exhaustion. After reaching exhaustion, participants completed a 2-minute cycling cool down at 60 rpm and 0 W. At the end of each 2-minute workload, participants' HR was recorded from Polar and they were asked to assess their level of physical exertion using the Borg Rating of Perceived Exertion (RPE) 10-point scale (69).

The second exercise participants completed was a resistance exercise circuit. This consisted of two identical circuits of 6 separate exercises completed for 8 repetitions each. The pace of the exercises was set at 3-0-1-0, meaning that the eccentric component of each exercise was performed for 3 seconds, followed by a pause, and then the concentric component of the exercise was performed for 1 second, followed by another pause. The 6 exercises performed were dumbbell bicep curls, Romanian deadlifts, Bulgarian split squat, dumbbell bench press, dumbbell shoulder press, and dumbbell step-ups. Prior to performing each exercise, participants were instructed to select a suitable dumbbell weight with which they could maintain proper form for 8 repetitions before reaching muscular fatigue. Each set of exercises was

separated by 60 seconds of rest (e.g. 8 repetitions of dumbbell bicep curls, 60 seconds of rest, 8 repetitions of Romanian deadlifts, 60 seconds of rest, etc.). After all 6 exercises had been completed once, participants had a 5-minute rest and then repeated the entire circuit of exercises again in the same order. After each exercise, participants' HR was recorded from Polar. After the first and last exercise of each circuit (i.e. dumbbell bicep curls and dumbbell step-ups), participants were asked to assess their level of physical exertion using the RPE scale.

After the VO_2 peak-T/ VO_2 peak-C and resistance exercises had been completed, participants left the laboratory.

4.1.2.2. Exercise Visit #2

After ≥ 24 hours, participants returned to the research laboratory in exercise clothing to complete ADL and HIIT-T/HIIT-C. The Cosmed K4b²/K5 and both activity monitors were placed on participants in the same locations and on the same wrists as the previous laboratory visit.

All participants were again given a 5-minute rest period for baseline assessment and recovery, before and after each exercise, respectively. Participants were instructed to remain seated during these assessments. Once the Cosmed K4b²/K5 and both activity monitors had been placed into the appropriate activity mode (Appendix A), study investigators started recording on all devices, which began the 5-minute, pre-exercise rest. Five minutes after each exercise ended, the devices were stopped and the 5-minute, post-exercise rest ended.

Participants performed the ADL first. Participants performed six ADL to simulate common, daily chores: sitting on a chair, washing dishes, sweeping the floor, organizing a room and moving light furniture, scrubbing the walls and floor, and self-paced ascending and descending of a flight of stairs. Each activity was performed for 3 minutes in duration, for a total time of 18 minutes. At the end of each activity, participants' HR was recorded from Polar and they were asked to assess their level of physical exertion using the RPE scale.

The second exercise participants performed was the HIIT-T/HIIT-C. Participants performed HIIT in the same mode of exercise (i.e. treadmill or cycle ergometer) as was used for the VO₂ peak test during the first laboratory visit. For HIIT-T, participants alternated between 6 low-intensity (30% of HRR) and 5 high-intensity (80% of HRR) intervals on the Spirit CT850 treadmill (70). Prior to the start of exercise, HRR was calculated for 30% and 80% intensity for each participant from HR measurements taken during the VO₂ peak test using the Karvonen formula (71):

$$\text{HRR} = [(\text{HR}_{\text{max}} - \text{HR}_{\text{rest}}) \times \% \text{ intensity}] + \text{HR}_{\text{rest}}$$

To begin, participants walked at a speed and incline corresponding to 30% of their HRR, as calculated by the Karvonen formula, for 2 minutes and then ran at a speed and incline corresponding to 80% of their HRR for 2 minutes. This continued until all 6 low-intensity and 5 high-intensity intervals had been completed. After completing the intervals, participants performed a 2-minute walking cool down at a self-selected speed and incline. For the HIIT-C, participants alternated between 6 low-intensity (30% of peak power output) and 5 high-intensity (80% of peak power output) intervals on the Lode Corival cycle ergometer. Prior to the start of exercise, power output was calculated

for 30% and 80% of peak power for each participant based on the peak power achieved during the VO_2 peak test. Peak power was multiplied by 0.3 and 0.8 to determine the power output for low- and high-intensity intervals, respectively. Participants cycled at a cadence of 60 rpm and 30% of peak power for 2 minutes and then cycled at a cadence of 60 rpm and 80% of peak power for 2 minutes. This continued until all 6 low-intensity and 5 high-intensity intervals had been completed. After completing the intervals, participants performed a 2-minute cycling cool down at a self-selected cadence and power output. At the end of each 2-minute workload, participants' HR was recorded from Polar and they were asked to assess their level of physical exertion using the RPE scale.

After the ADL and HIIT-T/HIIT-C had been completed, participants left the laboratory.

4.1.3. Devices

During all exercises, participants wore three research-grade devices (Cosmed K4b²/K5, Polar A300, Polar H7) and two wrist-worn, consumer-grade activity monitors (FC2, GVHR+).

4.1.3.1. Research-Grade Devices

The Cosmed K4b² and Cosmed K5 (Cosmed, Rome, Italy) are validated, wearable, and portable indirect calorimeters that measure various metabolic parameters, such as volume of oxygen consumption (VO_2), volume of carbon dioxide production (VCO_2), and ventilation (72–75). The Cosmed K4b²/K5 units were calibrated

before each testing session according to the manufacturer's instructions. The Cosmed K4b²/K5 calculated EE from the direct measurement of VO₂ and VCO₂. EE data was downloaded from the Cosmed K4b²/K5-associated software platform, OMNIA (version 1.6.5).

The Polar H7 (Polar Electro, Kempele, Finland) chest strap HR monitor measured HR at the chest through electrodes on a wearable chest strap and transmitted it to the Polar A300 (Polar Electro, Kempele, Finland) wrist watch via Bluetooth (76). The Polar A300 and H7 were used to monitor participants' HR during exercise. HR measurements from these devices were used to calculate HRR for the low- and high-intensity intervals in HIIT-T.

4.1.3.2. Consumer-Grade Devices

Although there are many consumer-grade activity monitors that are currently available, the FC2 and GVHR+ were chosen for this study based on budgetary considerations and their ability to integrate with an AP control system running on an Android platform.

The Fitbit Charge 2 (FC2) (2017 version, Fitbit, Inc., California, USA) is a multi-sensor activity monitor that has an accelerometer, altimeter, and built-in PPG sensor that uses the "PurePulse" wrist HR technology to measure HR at the wrist (77). The firmware version of the FC2 used in the study was 22.54.6. For the study, the FC2 recorded each exercise in the mode most closely-related to the exercise being performed (Appendix A). After each exercise session was recorded, data was uploaded from the FC2 to the Fitbit app (version 2.35) via Bluetooth. A new Fitbit user account

was created for each participant prior to the start of the study. These accounts were used to access participants' EE data after the study was completed. Fitbit only provided a total EE (kcal) value for each exercise session recorded. Fitabase (Small Steps Labs, California, USA) is a third-party research platform designed to collect data from Fitbit using the developer application programming interface (API). Fitabase provided access to data at a higher sample rate, including EE measured per minute, and access to METs estimates.

The Garmin vívosmart HR+ (GVHR+) (2016 version, Garmin International, Inc., Kansas, USA) is a multi-sensor activity monitor that has an accelerometer, altimeter, GPS, and built-in PPG sensor that uses the "Elevate" wrist HR technology to measure HR at the wrist (78). The firmware version of the GVHR+ used in the study was 3.20. For the study, the GVHR+ recorded each exercise in the mode most closely-related to the exercise being performed (Appendix A). After each exercise session was recorded, data was uploaded from the GVHR+ to the Garmin Connect app (version 3.17) via Bluetooth. A new Garmin user account was created for each participant prior to the start of the study. These accounts were used to access participants' EE data after the study was completed. Garmin provided a total EE (kcal) value for each exercise session recorded.

4.1.4. Statistical Analysis

The sample size of this study was calculated using the primary outcome measure of accuracy (i.e. %RE) in EE. The minimal difference in %RE to be detected was set as 5%, with a power of 0.8 and an alpha level of 0.05. Using G*Power (Heinrich Heine

University Düsseldorf, Düsseldorf, Germany, version 3.1.9.3), we calculated a sample size of 26 total subjects.

Data from the consumer-grade activity monitors were compared to the Cosmed K4b²/K5, which was used as the reference standard for EE, METs, and VO₂ measurements. Using the METs measurements from the FC2, VO₂ estimates were calculated using the following equation (79):

$$\text{VO}_2 \text{ (mL/min/Kg)} = \text{METs} \times 3.5$$

Mean relative percentage error (%RE) and mean absolute percentage error (MAPE) were calculated to determine the amount of error between the consumer-grade devices and the reference standard. Mean relative percentage error and MAPE were calculated using the following formulas (41):

$$\%RE = \frac{\sum [(\text{consumer device} - \text{research device}) / \text{research device}] \times 100}{n}$$

$$\text{MAPE} = \frac{\sum |[(\text{consumer device} - \text{research device}) / \text{research device}]| \times 100}{n}$$

No standardized thresholds exist for high or low %RE or MAPE, but a 10% mean error limit for the various activity was adopted for this study, similar to other studies (50,51,53). Error in EE was calculated across an entire exercise session as higher resolution data could not be obtained from all devices.

Mixed-effects analysis of variance (ANOVA) with the Geisser-Greenhouse correction and Tukey's honestly significant difference (HSD) post-hoc test was performed to assess the difference in total EE between the Cosmed K4b²/K5, FC2, and GVHR+ for each exercise. Matched paired t-tests were performed to assess the

difference in %RE and MAPE of EE between the FC2 and GVHR+ for each exercise. One-way ANOVA with Tukey's HSD post-hoc test was performed to assess the difference in %RE of EE, METs, and VO_2 and MAPE of EE between exercises within each device. Unpaired t-tests with Welch's correction were performed to assess the difference in %RE of EE between dominant and non-dominant hand wear for each device. Linear regression, the Pearson (r) correlation coefficient and Bland-Altman analysis were used to assess the agreement and bias between the consumer-grade devices and the reference standard. Linear regression and the Pearson (r) correlation coefficient were also used to examine the relationship between %RE and different anthropometric measures. All statistical analyses were conducted in GraphPad Prism 8 (GraphPad Software, California, USA, version 8.0.0). Statistical significance was set at $P < 0.05$. All data are presented as mean \pm SD.

4.2. RESULTS

4.2.1. Participants

All 20 participants completed the study. Table 10 describes participant characteristics.

4.2.2. Data Loss

There was no data loss from the FC2, but there was some data loss that occurred for the GVHR+. For two participants, study investigators failed to stop and restart recording on the GVHR+ between VO_2 peak and resistance exercise, so total EE for each individual activity was not known. One participant accidentally hit the stop

button on the GVHR+ during HIIT, so the second half of the exercise session was not recorded and the total EE for that activity was unknown. For unknown reasons, no EE data was recorded on the GVHR+ for any activity for two participants.

4.2.3. Energy Expenditure Accuracy

Both the FC2 and GVHR+ performed reasonably well at measuring EE overall, but considerable error was displayed during some activities, specifically during cycling activities for the FC2 and resistance exercise for the GVHR+. Table 11 presents pooled EE data from each exercise for the FC2 and the GVHR+.

The FC2 displayed a significant underestimation in total EE measurements during VO₂ peak (Cosmed: 231.3 ± 71.9 kcal, FC2: 162.9 ± 69.3 kcal, $P < 0.01$) and HIIT (Cosmed: 239.7 ± 58.6 kcal, FC2: 169.7 ± 71.2 , $P < 0.01$) compared to the Cosmed K4b²/K5, while the GVHR+ displayed a significant overestimation in total EE during VO₂ peak (FC2: 162.9 ± 69.3 kcal, GVHR+: 210.8 ± 49.7 kcal, $P < 0.05$) and resistance exercise (FC2: 130.2 ± 46.2 kcal, GVHR+: 179.8 ± 56.8 kcal, $P < 0.0001$) compared to the FC2 (Figure 1).

The correlational relationships in EE measurements between Cosmed and the consumer-grade devices ranged from -0.26 to 0.88 for the FC2 and 0.11 to 0.60 for the GVHR+ (Table 11). The FC2 had the strongest correlation with Cosmed during HIIT-T ($r = 0.88$) and the weakest correlation during HIIT-C ($r = -0.26$), while the GVHR+ also had the strongest correlation with Cosmed during HIIT-T ($r = 0.60$) and the weakest correlation during VO₂ peak-T ($r = 0.11$).

Correlation plots for total EE on Cosmed and total EE on the FC2 and GVHR+ demonstrated that both devices generally displayed an underestimation in EE compared to Cosmed, as seen by the majority of EE values appearing below the line of identity on the plots (Figure 2). The greatest underestimation in EE measurements occurred during VO₂ peak and HIIT on the FC2, with almost all of the EE values below the line of identity (Figure 2A and 2G), and the greatest overestimation occurred during resistance exercise on the GVHR+, with the majority of the EE values appearing above the line of identity (Figure 2D).

MAPE \pm SD for the FC2 and GVHR+ was $27.0 \pm 21.8\%$ and $25.1 \pm 17.3\%$, respectively (Figure 3). The lowest MAPE in EE was observed during all treadmill exercises ($16.9 \pm 10.9\%$) for the FC2 and ADL ($17.0 \pm 13.7\%$) for the GVHR+. The highest MAPE was observed during all cycling exercises ($42.7 \pm 26.8\%$) for the FC2 and resistance exercise ($35.7 \pm 19.7\%$) for the GVHR+. The GVHR+ was significantly more accurate than the FC2 at measuring EE during VO₂ peak (FC2: $31.5 \pm 21.5\%$, GVHR+: $22.9 \pm 16.8\%$, $P < 0.05$) and during all cycling exercises (FC2: $42.7 \pm 26.8\%$, GVHR+: $22.8 \pm 16.6\%$, $P < 0.05$).

Mean relative percentage error in EE measurements from the FC2 and GVHR+ differed significantly, with the GVHR+ displaying less negative bias (i.e. underestimation) overall (FC2: $-19.3 \pm 28.9\%$, GVHR+: $-1.6 \pm 30.6\%$, $P < 0.001$) (Figure 4). The lowest mean %RE in EE measurements was observed during ADL ($-8.8 \pm 29.2\%$) for the FC2 and VO₂ peak-C ($-4.5 \pm 25.3\%$) and HIIT-T ($-4.7 \pm 29.3\%$) for the GVHR+. The highest mean %RE was observed during VO₂ peak-C ($-39.1 \pm 30.6\%$) and HIIT-C ($-41.9 \pm 31.3\%$) for the FC2 and resistance ($21.0 \pm 35.7\%$) for the GVHR+.

Bland-Altman plots demonstrated that both the FC2 and GVHR+ displayed a negative bias in EE measurements when compared to Cosmed, with the FC2 displaying a greater mean difference from Cosmed than the GVHR+ (FC2: -43.1 ± 63.4 kcal, GVHR+: -12.2 ± 61.5 kcal) (Figure 5). However, the plots also revealed similar 95% LoA for both the FC2 (difference: 248.4 kcal) and the GVHR+ (difference: 241.0 kcal). In addition, the plots indicated that the greatest difference in EE measurements occurred during activities performed on the cycle ergometer, as demonstrated by the majority of EE values lying outside the LoA when cycling activities were performed.

Both the FC2 and GVHR+ demonstrated negative bias when activities were performed on the treadmill (FC2: $-15.1 \pm 13.5\%$, GVHR+: $-7.4 \pm 30.1\%$) (Figure 6). For activities performed on the cycle ergometer, both devices displayed a negative bias, but there was significantly greater mean error on the FC2 compared to the GVHR+ (FC2: $-40.5 \pm 30.2\%$, GVHR+: $-7.9 \pm 27.6\%$, $P < 0.0001$).

Correlation plots of EE measurements in kcal/min from Cosmed and the FC2 demonstrate that the FC2 primarily underestimated EE during VO_2 peak, HIIT, all pooled treadmill exercises, and all pooled cycle ergometer exercises compared to Cosmed, as seen by the majority of EE values appearing below the line of identity (Figure 7). The FC2 displayed the greatest accuracy in EE measurements during ADL (Figure 7C), while the greatest inaccuracy and underestimation in EE measurements occurred during cycling exercises (Figure 6F). The EE values clustered in a horizontal line above the x-axis in Figure 6F demonstrates that although Cosmed was measuring EE from 1-20 kcal/min, the FC2 was unable to accurately detect these values and

inaccurately measured EE as being between only 1-2 kcal/min during 49.5% of cycling activities.

Individual subject data from cycling exercise demonstrates that the underestimation in EE measurements occurred for at least 3 minutes or more during all VO₂ peak-C and HIIT-C sessions by the FC2 and low, inaccurate EE measurements were reported during partial and for full durations of activities (Appendix B). In some instances, the FC2 reported low, inaccurate EE values (<2 kcal/min) for the first 3-6 minutes of activity, but then reported higher values (>2 kcal/min) for the rest of the activity (Appendix BA, BC, BF, and BJ). In other cases, the FC2 reported low, inaccurate EE values for the majority or full duration of exercise (Appendix BE, BG, BH, and BI).

There was no significant difference in the %RE of EE measurements between dominant hand and non-dominant hand wear for either the FC2 or the GVHR+ (Appendix C).

Moderate or strong relationships were not displayed between %RE for the devices and most of the anthropometrics measured (Appendix D). However, a moderate, positive relationship was displayed between %RE in EE for the GVHR+ and BF% ($r=0.40$, $P<0.05$) (Appendix DJ).

4.2.4. METs and VO₂ Accuracy

METs values for the FC2 were retrieved from Fitabase and used to calculate VO₂. METs and VO₂ measurements from the FC2 displayed a negative bias (METs: $-14.6 \pm 30.7\%$, VO₂: $-14.5 \pm 30.0\%$) (Appendix E), similar to EE measurements. The

Garmin Connect application does not provide METs data, so METs and VO_2 measurements could not be analyzed for the GVHR+.

5. STUDY B: PILOT STUDY

5.1. METHODS

5.1.1. Participants

Fifteen adults (9 T1D, 6 healthy controls) were recruited for this study. All participants provided written informed consent before taking part in the study. All participants were screened for cardiovascular complications using the Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) (64). Physical activity levels were assessed in all participants using the International Physical Activity Questionnaire (IPAQ) (65). This study was approved by the Human Participants Review Sub-Committee at York University (Toronto, Ontario) (Certificate #: e2018-010).

5.1.2. Study Protocol

Participants attended the research laboratory on two separate occasions, separated by a minimum of 24 hours. Anthropometrics were gathered on all participants at the beginning of their first laboratory visit. After completing this, participants performed a progressive aerobic exercise test to peak oxygen consumption on a treadmill (VO₂ peak-T), a resistance exercise circuit, and high-intensity interval training on a cycle ergometer (HIIT-C) (described in more detail in section 5.1.2.1. *Exercise Visit #1*). During the second laboratory visit, participants performed a series of common activities of daily living (ADL) and high-intensity interval training on a treadmill (HIIT-T) (described in more detail in section 5.1.2.2. *Exercise Visit #2*). All participants wore a portable metabolic system (Cosmed K5), a HR monitor (Polar H10), and two wrist-worn,

consumer-grade activity monitors (FI, GV3) during all exercises (described in more detail in section 5.1.3. *Devices*).

During the first laboratory visit and prior to starting exercise, anthropometrics were gathered on all participants. Height was measured using a wall-mounted stadiometer to the nearest 0.1 cm. Participants were instructed to remove their shoes and stand up straight against the stadiometer with their back and heels touching the wall. They were instructed to take a deep breath and exhale, after which point an investigator lowered the headpiece to the top of the participant's head. The participant was asked to step away from the wall and their height was recorded. Weight and BF% were measured using the Tanita BF-350 Total Body Composition Analyzer scale (Tanita Corporation, Tokyo, Japan) to the nearest 0.1 kg and 0.1%, respectively. Participant's sex, age, and height were entered into the Tanita scale. Once this information had been entered into the scale, participants were instructed to remove their shoes and socks, step on the scale, and remain standing on the scale until their weight and BF% were displayed on the screen and recorded. WC and wrist circumference were measured using a soft vinyl tape measure by one of the study investigators to the nearest 0.1 cm. WC was measured horizontally around the abdomen at the level of the iliac crest (66). Wrist circumference was measured adjacent and proximal to the styloid bone of the ulna on the dominant hand. Blood pressure and resting HR were measured using the OMRON M7 Intelli IT blood pressure monitor and OMRON Intelli Wrap HEM-FL31 blood pressure cuff (OMRON Corporation, Kyoto, Japan). Participants were seated in a chair with the blood pressure cuff wrapped around their left arm, which was placed on a table at chest height next to them. Participants were instructed to relax, not speak, and keep

their legs uncrossed and feet flat on the floor during the measurements (67,68). Duplicate blood pressure and resting HR measurements were taken 60 seconds apart and then averaged together. Skin tone measurements were taken at the wrist independently by two investigators using the Von Luschan Chromatic scale (1-36 tile scale) (80). Measurements were always taken away from windows and under florescent lights (80). The average of the two measurements were converted to the Fitzpatrick skin tone scale (1-6 tile scale) (46,81,82).

5.1.2.1. Exercise Visit #1

Participants arrived at the lab in exercise-appropriate clothing (e.g. t-shirt, tank top, shorts, leggings, etc.). After initial anthropometrics had been measured and recorded, two consumer-grade activity monitors (FI, GV3) were placed on opposing wrists (as per manufacturer's instructions) on the participant in a randomized and counterbalanced method. In addition, all participants wore a portable metabolic system (Cosmed K5) and a HR monitor (Polar H10) during all exercises.

All participants were given a 5-minute rest period for baseline assessment and recovery, before and after each exercise, respectively. Participants were instructed to remain seated during these assessments. Once the Cosmed K5 and both activity monitors had been placed into the appropriate activity mode (Appendix F), study investigators started recording on all devices, which began the 5-minute, pre-exercise rest. Five minutes after each exercise ended, the devices were stopped and the 5-minute, post-exercise rest ended.

The first exercise participants performed was the VO₂ peak-T. Each participant began with a 4-minute walking warm-up (3.0 mph, 0% grade for 2 min and then 3.0 mph, 5% grade for 2 min) on the Spirit CT850 treadmill (Dyaco International, Taipei, Taiwan). After the warm-up, participants self-selected a comfortable running speed between 4 to 6 mph, and subsequently, the treadmill incline was increased by 2% every 2 minutes until the participants reached volitional exhaustion (52). After reaching exhaustion, participants completed a 2-minute walking cool down at 2.5 mph and 0% incline. At the end of each 2-minute workload, participants' HR was recorded from Polar and they were asked to assess their level of physical exertion using the RPE scale.

The second exercise participants completed was a resistance exercise circuit. This consisted of two identical circuits of 6 separate exercises completed for 8 repetitions each. The pace of the exercises was set at 3-0-1-0, meaning that the eccentric component of each exercise was performed for 3 seconds, followed by a pause, and then the concentric component of the exercise was performed for 1 second, followed by another pause. The 6 exercises performed were dumbbell bicep curls, Romanian deadlifts, Bulgarian split squat, push-ups, dumbbell shoulder press, and dumbbell step-ups. Prior to performing each exercise, participants were instructed to select a suitable dumbbell weight with which they could maintain proper form for 8 repetitions before reaching muscular fatigue. Dumbbells were used for all exercises, except for push-ups, which were performed using body weight. Each set of exercises was separated by 60 seconds of rest (e.g. 8 repetitions of dumbbell bicep curls, 60 seconds of rest, 8 repetitions of Romanian deadlifts, 60 seconds of rest, etc.). After all 6 exercises had been completed once, participants had a 5-minute rest and then repeated

the entire circuit of exercises again in the same order. After each exercise, participants' HR was recorded from Polar. After the first and last exercise of each circuit (i.e. dumbbell bicep curls and dumbbell step-ups), participants were asked to assess their level of physical exertion using the RPE scale.

Following a 1-hour rest, participants performed HIIT-C. Participants alternated between 6 low-intensity (RPE of 2) and 5 high-intensity (RPE of 8) intervals on the Lode Corival cycle ergometer (Lode BV, Groningen, Netherlands) (70). Participants were instructed to cycle at a power output that they considered to be equivalent to an RPE of ~2 and ~8 for the low- and high-intensity intervals, respectively. Once the power outputs had been determined for each interval during the first low- and first high-intensity intervals, they were kept consistent for the rest of the exercise. Participants cycled at a cadence of 60 rpm and a power output equivalent to an RPE of 2 for 2 minutes and then cycled at a cadence of 60 rpm and a power output equivalent to an RPE of 8 for 2 minutes. This continued until all 6 low-intensity and 5 high-intensity intervals had been completed. After completing the intervals, participants performed a 2-minute cycling cool down at a self-selected cadence and power output. At the end of each 2-minute workload, participants' HR was recorded from Polar and they were asked to assess their level of physical exertion using the RPE scale.

After the VO_2 peak-T, resistance exercises, and HIIT-C had been completed, participants left the laboratory.

5.1.2.2. Exercise Visit #2

After ≥ 24 hours, participants returned to the research laboratory in exercise clothing to complete ADL and HIIT-T. The Cosmed K5 and both activity monitors were placed on participants in the same locations and on the same wrists as the previous laboratory visit.

All participants were again given a 5-minute rest period for baseline assessment and recovery, before and after each exercise, respectively. Participants were instructed to remain seated during these assessments. Once the Cosmed K5 and both activity monitors had been placed into the appropriate activity mode (Appendix F), study investigators started recording on all devices, which began the 5-minute, pre-exercise rest. Five minutes after each exercise ended, the devices were stopped and the 5-minute, post-exercise rest ended.

Participants performed the ADL first. As in Study A, participants performed six ADL to simulate common, daily chores: sitting on a chair, washing dishes, sweeping the floor, organizing a room and moving light furniture, scrubbing the walls and floor, and self-paced ascending and descending of a flight of stairs. Each activity was performed for 3 minutes in duration, for a total time of 18 minutes. At the end of each activity, participants' HR was recorded from Polar and they were asked to assess their level of physical exertion using the RPE scale.

The second exercise participants performed was the HIIT-T. Participants alternated between 6 low-intensity (30% of HRR) and 5 high-intensity (80% of HRR) intervals on the Spirit CT850 treadmill (70). Prior to the start of exercise, HRR was

calculated for 30% and 80% intensity for each participant from HR measurements taken during the VO₂ peak test using the Karvonen formula (71):

$$\text{HRR} = [(\text{HR}_{\text{max}} - \text{HR}_{\text{rest}}) \times \% \text{ intensity}] + \text{HR}_{\text{rest}}$$

To begin, participants walked at a speed and incline corresponding to 30% of their HRR, as calculated by the Karvonen formula, for 2 minutes and then ran at a speed and incline corresponding to 80% of their HRR for 2 minutes. This continued until all 6 low-intensity and 5 high-intensity intervals had been completed. After completing the intervals, participants performed a 2-minute walking cool down at a self-selected speed and incline. At the end of each 2-minute workload, participants' HR was recorded from Polar and they were asked to assess their level of physical exertion using the RPE scale.

After the ADL and HIIT-T had been completed, participants left the laboratory.

5.1.3. Devices

During all exercises, participants wore two research-grade devices (Cosmed K5, Polar H10) and two wrist-worn, consumer-grade activity monitors (FI, GV3).

5.1.3.1. Research-Grade Devices

The Cosmed K5 (Cosmed, Rome, Italy) is a validated, wearable, and portable indirect calorimeter that measures various metabolic parameters, such as VO₂, VCO₂, and ventilation (74,75). The Cosmed K5 unit was calibrated before each testing session according to the manufacturer's instructions. The Cosmed K5 calculated EE from the

direct measurement of VO_2 and VCO_2 . EE data was downloaded from the Cosmed K5-associated software platform, OMNIA (version 1.6.5).

The Polar H10 (Polar Electro, Kempele, Finland) chest strap HR monitor measured HR at the chest through electrodes on a wearable chest strap and transmitted it to the Polar Beat app (version 2.6.3) via Bluetooth (76). The Polar H10 and the Polar Beat app were used to monitor participants' HR during exercise. HR measurements from Polar were used to calculate HRR for the low- and high-intensity intervals in HIIT-T.

5.1.3.2. Consumer-Grade Devices

Based on the greater accuracy displayed by the GVHR+ in Study A, the GV3 was selected to be used in this study, which is the newest version of the GVHR+. Due to the poor results displayed by the FC2 in Study A, the FC2 was not chosen to be used in this study, but the FI, a newer and higher end activity monitor developed by Fitbit, was chosen to be used in this study instead. Both devices have the ability to integrate with an AP control system running on an Android platform.

The Fitbit Ionic (FI) (2018 version, Fitbit, Inc., California, USA) is a multi-sensor activity monitor that has an accelerometer, altimeter, gyroscope, GPS, and built-in PPG sensor that uses the "PurePulse" wrist HR technology to measure HR at the wrist (83). The firmware version of the FI used in the study was 32.10.20. For the study, the FI recorded each exercise in the mode most closely-related to the exercise being performed (Appendix F). After each exercise session was recorded, data was uploaded from the FI to the Fitbit app (version 2.81.1) via Bluetooth. A new Fitbit user account

was created for each participant prior to the start of the study. These accounts were used to access participants' EE data after the study was completed. Fitbit provided a total EE (kcal) value for each exercise session recorded. Due to limited funding, a Fitabase subscription could not be acquired for this study, so data at a higher sample rate (i.e. EE in kcal/min) and METs data could not be analyzed in this study.

The Garmin vívosmart 3 (GV3) (2018 version, Garmin International, Inc., Kansas, USA) is a multi-sensor activity monitor that has an accelerometer, altimeter, and built-in PPG sensor that uses the “Elevate” wrist HR technology to measure HR at the wrist (84). The firmware version of the GV3 used in the study was 5.00. For the study, the GV3 recorded each exercise in the mode most closely-related to the exercise being performed (Appendix F). After each exercise session was recorded, data was uploaded from the GV3 to the Garmin Connect app (version 4.15.1.2) via Bluetooth. A new Garmin user account was created for each participant prior to the start of the study. These accounts were used to access participants' EE data after the study was completed. Garmin provided a total EE (kcal) value for each exercise session recorded.

5.1.4. Glucose

Glucose was monitored before, during, and after exercise using SMBG and CGM. SMBG was performed before and after each exercise, while CGM was worn continuously. Sixty seconds before the start of exercise (during the 5-minute, pre-exercise rest) and immediately after completing exercise (during the 5-minute, post-exercise rest), duplicate SMBG measurements were taken. Participants were instructed to clean their finger with a BD alcohol swab (Becton, Dickinson and Company, New

Jersey, USA) and then dry their finger with a tissue. They were asked to squeeze and wipe away the first drop of blood to avoid testing blood that may be contaminated with substances from the surface of the skin, such as alcohol, food, or sweat, that may impact the accuracy of the BG measurement (85,86). Participants were asked to produce a second and third drop of blood, which were then tested using the CONTOUR NEXT LINK glucometer and CONTOUR NEXT blood glucose test strips (Ascensia Diabetes Care Holdings AG, Basel, Switzerland) (87). The two BG measurements were averaged together. If the measurements were >1.0 mmol/L apart, a third measurement was taken and the two closest BG measurements were averaged together.

BG was checked prior to the start of exercise to ensure individuals with T1D were in a safe glycemic range (i.e. 5-15 mmol/L) to exercise (88). If BG from SMBG was <5 mmol/L prior to exercise start, participants were asked to ingest 3-4 Dex4 glucose tablets (12-16g of carbohydrates) (AMG Medical Inc., Quebec, Canada) and wait 15 minutes to allow BG to rise to a safe level. After 15 minutes, BG was checked using SMBG. If BG was ≥ 5 mmol/L, exercise was initiated. If BG was still below <5 mmol/L, participants were asked to ingest another 1-2 Dex4 tabs (4-8 g of carbohydrates) and wait another 15 minutes. If BG was >15 mmol/L, participants were tested for ketones using the FreeStyle Precision Neo meter and FreeStyle Precision Blood β -ketone test strips (Abbott Laboratories, Illinois, USA). If ketones were <1.5 mmol/L, exercise was initiated. If ketones were ≥ 1.5 mmol/L, exercise was not started and participants were asked to initiate their usual diabetes care (88).

At least ≥ 24 hours prior to their first laboratory visit, participants were set up with a Dexcom G5 CGM system (Dexcom, Inc., California, USA). The Dexcom G5 CGM

measures interstitial glucose every 5 minutes through a sensor inserted into subcutaneous tissue and transmits it to the Dexcom G5 Mobile app (version 1.7.5) on the user's smartphone via Bluetooth (89–91). Participants' glucose data from exercise and for 12 hours post-exercise was downloaded using Dexcom CLARITY (version 3.22.1), a web-based software for healthcare professionals and researchers to access and analyze CGM data.

Participants' BG was closely monitored throughout exercise using CGM. If CGM indicated glucose was exceeding the safe glycemic range for exercise, BG was tested using SMBG. If participants' BG was <3.9 mmol/L, exercise was stopped and participants were asked to ingest 3-4 Dex4 tabs and wait 15 minutes. After 15 minutes, BG was checked using SMBG. If BG was ≥ 3.9 mmol/L, exercise was continued. If BG was still below <3.9 mmol/L, participants were asked to ingest another 1-2 Dex4 tabs and wait another 15 minutes. If BG was >15 mmol/L, participants were tested for ketones. If ketones were <1.5 mmol/L, exercise was continued. If ketones were ≥ 1.5 mmol/L, exercise was terminated and participants were asked to initiate their usual diabetes care (88).

5.1.5. Statistical Analysis

The sample size of this study was calculated using the primary outcome measure of accuracy (i.e. %RE) in EE. The minimal difference in %RE to be detected was set as 5%, with a power of 0.8 and an alpha level of 0.05. Using G*Power (Heinrich Heine University Düsseldorf, Düsseldorf, Germany, version 3.1.9.3), we calculated a sample size of 198 total subjects.

Data from the consumer-grade activity monitors were compared to the Cosmed K5, which was used as the reference standard for EE measurements. Mean relative percentage error (%RE) and MAPE were calculated to determine the amount of error between the consumer-grade devices and the reference standard. Mean relative percentage error and MAPE were calculated using the following formulas (41):

$$\%RE = \frac{\sum [(consumer\ device - research\ device) / research\ device] \times 100}{n}$$

$$MAPE = \frac{\sum | [(consumer\ device - research\ device) / research\ device] | \times 100}{n}$$

No standardized thresholds exist for high or low %RE or MAPE, but a 10% mean error limit for the various activity was adopted for this study, similar to other studies (50,51,53). Error in EE was calculated across an entire exercise session as higher resolution data could not be obtained from the devices.

CGM data was analyzed for percentage of time in hypoglycemia, euglycemia, and hyperglycemia during exercise and for 12 hours post-exercise. Based on the recommendations by Danne et al. (3), the euglycemic target range was set as 3.9-10.0 mmol/L, while <3.9 mmol/L was considered hypoglycemia and >10.0 mmol/L was considered hyperglycemia.

Mixed-effects analysis of variance (ANOVA) with the Geisser-Greenhouse correction and Tukey's HSD post-hoc test was performed to assess the difference in total EE between the Cosmed K5, FI, and GV3 for each exercise. Matched paired t-tests were performed to assess the difference in %RE and MAPE of EE between the FI and GV3 for each exercise. One-way repeated measures ANOVA with Tukey's HSD

post-hoc test was performed to assess the difference in %RE and MAPE of EE between exercises within each device. Unpaired t-tests with Welch's correction were performed to assess the difference in %RE of EE between T1D and control subjects and between dominant and non-dominant hand wear for each device. Linear regression, the Pearson (r) correlation coefficient and Bland-Altman analysis were used to assess the agreement and bias in total EE between the consumer-grade devices and the reference standard. Linear regression and the Pearson (r) correlation coefficient were also used to examine the relationship between %RE and different anthropometric measures. For SMBG measurements, matched paired t-tests were performed to assess the difference in BG from pre- to post-exercise in each visit. Unpaired t-tests with Welch's correction were performed to assess the difference in BG between T1D and control subjects at the start and end of each visit. Unpaired t-tests with Welch's correction were also performed to assess the change in BG during each exercise between T1D and control subjects. One-way repeated measures ANOVA with Tukey's HSD post-hoc test was performed to assess the change in BG between exercises within each group of subjects. All statistical analyses were conducted in GraphPad Prism 8 (GraphPad Software, California, USA, version 8.0.0). Statistical significance was set at $P < 0.05$. All data are presented as mean \pm SD.

5.2. RESULTS

5.2.1. Participants

Fourteen participants (8 T1D, 6 healthy) completed both visit 1 and 2. An additional participant with T1D only completed visit 1. All data collected, including the

participant who only completed visit 1, has been included in the Results. Table 12 describes participant characteristics.

5.2.2. Data Loss

There was some data loss that occurred for both the FI and GV3. The FI would not pair with one participant's Fitbit account prior to the start of VO₂ peak, so no EE data was recorded for this activity session. One participant did not complete HIIT-C, as they were the first subject to complete the study before the protocol was finalized. One other participant accidentally hit the stop button on the FI during HIIT-T, so the second half of the exercise session was not recorded and the total EE for that activity was unknown. Despite recording the exercise sessions, the GV3 did not report any total EE value during resistance exercise and HIIT-C for one participant and during VO₂ peak, resistance, and HIIT-C for another participant for unknown reasons.

During the data collection process, the Cosmed K5 began malfunctioning and was sent out for repair. The Cosmed Fitmate Pro (Cosmed, Rome, Italy) was used with two participants in place of the Cosmed K5, while it was being repaired. One participant completed visit 1 and 2 using the Cosmed Fitmate Pro, while the other only completed visit 1 while using it. While analyzing the data from the Cosmed K5 and Cosmed Fitmate Pro, it was determined that the EE measurements reported by the Cosmed Fitmate Pro were significantly different than the values reported by the Cosmed K5. When EE measurements from the activity monitors were compared to the Cosmed Fitmate Pro, they displayed significantly higher %RE and MAPE in comparison to those

that were compared to the Cosmed K5. Therefore, all data that was collected and compared to the Cosmed Fitmate Pro was excluded from the Results.

Some data loss also occurred for the Dexcom G5 CGM, as was expected based on our prior experience with this technology (sensor, transmitter, and receiver Bluetooth communication loss sometimes occurs). One participant, who only completed visit 1, did not provide access to their glucose data, so their glucose data could not be analyzed. Two participants experienced issues with the glucose sensor and transmitter properly reporting glucose data to the Dexcom G5 app on their smartphone, which resulted in a lot of missing data. For one participant, there was missing glucose data during and for 12 hours post-exercise for visit 1 and for 12 hours post-exercise for visit 2, while there was missing data for another participant during the 12 hours post-exercise following visit 2. Consequently, due to all the missing data, the glucose data for these subjects during these periods could not be analyzed and was excluded from the Results. In addition, the glucose sensor that was inserted into the skin became loose and came out of the skin in two participants by the end of visit 2. As a result, no glucose data was collected for these participants for the 12 hours post-exercise following visit 2.

5.2.3. Energy Expenditure Accuracy

Both the FI and GV3 performed reasonably well at measuring EE overall, but a considerable positive bias (i.e. overestimation) was displayed during resistance exercise on both devices. Table 13 presents pooled EE data from each exercise for the FI and the GV3.

The FI displayed a significant overestimation in total EE measurements during resistance exercise (Cosmed: 164.2 ± 47.0 kcal, FI: 249.3 ± 93.6 kcal, $P < 0.01$) and a significant underestimation in EE during HIIT-T (Cosmed: 257.8 ± 51.3 kcal, FI: 231.5 ± 57.7 , $P < 0.05$) compared to the Cosmed K5 (Figure 8).

The correlational relationships in EE measurements between Cosmed and the consumer-grade devices ranged from 0.58 to 0.92 for the FI and 0.04 to 0.89 for the GV3 (Table 13). The FI had the strongest correlation with Cosmed during VO₂ peak-T ($r = 0.92$) and the weakest correlation during HIIT-C ($r = 0.58$), while the GV3 also had the strongest correlation with Cosmed during VO₂ peak-T ($r = 0.89$) and the weakest correlation during resistance exercise ($r = 0.04$).

Correlation plots for total EE on Cosmed and total EE on the FI and GV3 demonstrated an overestimation in EE measurements on the FI during resistance exercise in both T1D and control subjects (Figure 9C). In addition, T1D subjects displayed an overestimation in EE measurements during both resistance exercise and HIIT-C on the GV3 compared to the control subjects, as seen by the separation of each group's EE values from the other on the plots (Figure 9D and 9J).

MAPE \pm SD for the FI and GV3 was $27.5 \pm 27.9\%$ and $27.0 \pm 25.0\%$, respectively (Figure 10). The lowest MAPE in EE was observed during HIIT-T ($11.0 \pm 6.6\%$) for the FI and VO₂ peak ($16.4 \pm 11.2\%$) for the GV3. The highest MAPE was observed during resistance exercise for both devices (FI: $56.1 \pm 44.4\%$, GV3: $44.6 \pm 34.3\%$).

Mean relative percentage error displayed a positive bias in EE measurements for the FI and GV3 compared to Cosmed, with the FI displaying a greater positive bias overall (FI: $15.7 \pm 36.0\%$, GV3: $7.0 \pm 36.3\%$) (Figure 11). The lowest %RE in measuring

EE was observed during VO₂ peak ($7.4 \pm 22.6\%$), HIIT-T ($-7.8 \pm 10.5\%$), and HIIT-C ($6.8 \pm 32.9\%$) for the FI and HIIT-T ($-2.2 \pm 26.1\%$) for the GV3. The highest %RE was observed during resistance exercise for both devices (FI: $56.1 \pm 44.4\%$, GV3: $28.6 \pm 49.7\%$). There was a significant difference in %RE of EE measurements during VO₂ peak between the FI and GV3 (FI: $7.4 \pm 22.6\%$, GV3: $-4.6 \pm 19.9\%$, $P < 0.05$).

Bland-Altman plots demonstrated a positive bias in EE measurements on the FI with a mean difference of 19.8 ± 65.1 kcal, while the GV3 displayed only a small positive bias in EE with a mean difference of 1.9 ± 72.0 kcal when compared to Cosmed (Figure 12). However, the plots revealed narrower 95% LoA for the FI (difference: 255.4 kcal) compared to the GV3 (difference: 282.4 kcal). In addition, the plots indicated that the greatest difference in EE measurements on the FI occurred during resistance exercise, as demonstrated by almost all of the EE values lying outside the LoA when resistance exercise was performed (Figure 12A).

Both the FI and GV3 demonstrated a negative bias for exercise performed on the treadmill (FI: $-0.2 \pm 18.9\%$, GV3: $-3.4 \pm 22.9\%$) and a positive bias for exercise performed on the cycle ergometer (FI: $6.8 \pm 32.9\%$, GV3: $5.8 \pm 43.1\%$) (Figure 13). However, there was no significant difference between the FI and GV3 in %RE of EE during treadmill and cycling exercises.

A significant difference in %RE in EE was displayed by the GV3 between T1D and control subjects (T1D: $22.6 \pm 35.4\%$; Control: $-15.6 \pm 24.0\%$, $P < 0.0001$), but no such difference was exhibited by the FI (T1D: $13.5 \pm 35.8\%$; Control: $19.5 \pm 37.0\%$) (Figure 14). The FI displayed a positive bias in %RE in EE in both T1D and control

subjects, while the GV3 displayed a positive bias in individuals with T1D and a negative bias in control subjects.

The GV3 demonstrated a positive bias in EE for individuals with T1D during all 5 exercises performed compared to control subjects, with a significant difference between groups during resistance exercise (T1D: $58.5 \pm 37.7\%$, Control: $-7.4 \pm 37.8\%$, $P < 0.05$), ADL (T1D: $23.3 \pm 37.8\%$, Control: $-12.0 \pm 14.4\%$, $P < 0.05$), and HIIT-C (T1D: $32.7 \pm 32.5\%$, Control: $-34.6 \pm 13.8\%$, $P < 0.01$) (Appendix G).

There were no significant differences in most anthropometric measures between T1D and control subjects, albeit our sample size was small, with the exception of the Fitzpatrick skin tone score and BF%. Individuals with T1D had significantly higher mean scores on the Fitzpatrick skin tone scale (T1D: 2.6 ± 0.5 , Control: 2.0 ± 0.0 , $P < 0.05$) and higher BF% (T1D: 26.9 ± 9.0 , Control: 17.7 ± 5.1 , $P < 0.05$) compared to control subjects (Appendix H).

There was also no significant difference in the %RE of EE measurements between dominant hand and non-dominant hand wear for either the FI or the GV3 (Appendix I).

Moderate or strong relationships were not displayed between %RE for the devices and most of the anthropometrics measured (Appendix J). However, a moderate, positive relationship was displayed between %RE in EE for the GV3 and BF% ($r = 0.52$, $P < 0.0001$) (Appendix JJ) and a moderate, negative relationship was displayed between %RE for the GV3 and VO_2 peak ($r = -0.40$, $P < 0.01$) (Appendix JX).

5.2.4. Glucose

Blood glucose and interstitial glucose concentrations were monitored in all individuals with T1D before, during, and for 12-hours post-exercise using SMBG and CGM, respectively. Two control subjects also agreed to monitor their glucose before, during, and immediately after exercise using SMBG and CGM. Only one of those control subjects agreed to wear a CGM system for 12-hours post-exercise to monitor their interstitial glucose.

Both T1D and control subjects displayed a decrease in mean BG from pre- to post-exercise during both visits, with the greatest change occurring in individuals with T1D during visit 1 (Figure 15). A mean decrease of 2.6 ± 3.8 mmol/L in BG was observed in individuals with T1D from pre- to post-exercise (Pre: 9.7 ± 3.2 mmol/L, Post: 7.2 ± 2.0 mmol/L, n=9) in visit 1, while a mean decrease of 1.3 ± 2.8 mmol/L was observed from pre- to post-exercise (Pre: 9.0 ± 1.8 mmol/L, Post: 7.7 ± 2.2 mmol/L, n=8) in visit 2. A mean decrease of 0.6 ± 0.1 and 1.6 ± 0.3 mmol/L was observed in the control subjects from pre- to post-exercise in visit 1 (Pre: 6.3 ± 0.8 mmol/L, Post: 5.7 ± 0.7 mmol/L, n=2) and in visit 2 (Pre: 6.9 ± 0.1 mmol/L, Post: 5.3 ± 0.4 mmol/L, n=2), respectively.

For individual exercises, the largest mean change in BG occurred during HIIT-C for individuals with T1D and during ADL for control subjects (Figure 16). In visit 1, individuals with T1D exhibited a mean increase of 0.9 ± 2.1 mmol/L in BG during the VO₂ peak test, followed by a mean decrease of 0.8 ± 2.0 and 2.1 ± 2.5 mmol/L in BG during resistance exercise and HIIT-C, respectively. Control subjects displayed a mean change in BG of 1.3 ± 2.0 mmol/L during VO₂ peak, -0.3 ± 0.2 mmol/L during resistance

exercise, and 0.5 ± 1.3 mmol/L during HIIT-C. In visit 2, a mean decrease of 0.1 ± 0.9 and 0.8 ± 1.6 mmol/L in BG was observed in individuals with T1D during ADL and HIIT-T, respectively, while control subjects displayed a mean decrease in BG of 2.6 ± 0.6 mmol/L during ADL and 0.4 ± 0.1 mmol/L during HIIT-T. A maximum change in BG of 7.3 mmol/L was observed in the T1D subjects during exercise, while only a maximum change of 3.0 mmol/L was observed in the control subjects.

Based on interstitial glucose levels as measured by CGM, individuals with T1D spent more time in euglycemia during and following visit 2 compared to visit 1, while control subjects were in euglycemia almost 100% of the time during and following both visits (Figure 17). Mean percent time in euglycemia during exercise was 54% and 86%, time in hypoglycemia was 6% and 0%, and time in hyperglycemia was 40% and 14% in individuals with T1D during visit 1 and 2, respectively. Mean percent time in euglycemia during exercise for control subjects was 100% in both visits. For 12 hours post-exercise, mean percent time in euglycemia was 59% and 66%, time in hypoglycemia was 6% and 3%, and time in hyperglycemia was 34% and 31% in individuals with T1D following visit 1 and 2, respectively. One control subject agreed to wear the CGM system for 12 hours post-exercise and their percent time in euglycemia was 100% for 12 hours post-exercise after visit 1 and 92% after visit 2, with the remaining 8% of time spent in hypoglycemia.

Following visit 1, the glucose levels in most individuals with T1D remained relatively stable and within the glucose target range (i.e. 3.9-10.0 mmol/L) for these individuals for ~3-4 hours post-exercise. After 4 hours, mean interstitial glucose exceeded 10 mmol/L and remained elevated in the hyperglycemic range for 2 hours

before returning to and staying in euglycemia for the next 6 hours. Following visit 2, mean interstitial glucose remained in target for 5.5 hours. Mean interstitial glucose entered hyperglycemia for ~1.5 hours, returned to euglycemia for ~1 hour, before returning to and staying in a hyperglycemic range for the next 5 hours, although there was large variability between subjects during this 5-hour period. For one control subject, interstitial glucose remained in euglycemia for the entire 12 hours post-exercise following visit 1. Following visit 2, interstitial glucose for the control subject was in euglycemia for ~2 hours, before entering hypoglycemia for ~1 hour, and then returning to and remaining stable in euglycemia for the next 9 hours (Figure 18).

6. DISCUSSION

6.1. PRINCIPAL FINDINGS

This project examined the accuracy of four wrist-worn, consumer-grade activity monitors at measuring EE during a variety of physical activities. Similar to previous studies (27,34,38,40,41,50,51,53), all four activity monitors tested displayed a high overall MAPE in EE measurements, ranging from 25-28%. The GVHR+ (-1.6%) and the GV3 (7.0%) were the only two devices to display a %RE of $\leq 10\%$, while the FC2 (-19.3%) and FI (15.7%) displayed a %RE of $>10\%$. The FC2 displayed a significant underestimation in EE measurements during cycling, while the FI, GVHR+, and GV3 all displayed an overestimation in EE during resistance exercise. The GV3 significantly overestimated EE in individuals with T1D compared to individuals without T1D. A greater mean decrease in BG was observed in individuals with T1D during exercise visit 1 compared to visit 2. There was a mean decrease in BG in individuals with T1D during all activities, except during the VO_2 peak test where they displayed a mean increase in BG.

The FC2 displayed a significantly higher MAPE (42.7%) during cycling than any other device tested and a significant underestimation in EE measurements (-40.5%) during cycling activities compared to other activities, which was not exhibited in the other three activity monitors assessed. Boudreaux et al. (53) observed the highest MAPE (75%) for the FC2 out of seven devices tested during cycling with an underestimation in EE in 80% of subjects. They also assessed the Garmin vivosmart HR, an earlier model to the GVHR+, which displayed no significant difference in EE measurements compared to indirect calorimetry, similar to our findings (53).

The FI (56.1%), GVHR+ (21.0%), and GV3 (28.6%) all displayed a greater %RE and greater overestimation in EE during resistance exercise compared to all other activities. These findings were similar to the findings by Boudreaux et al. (53), who observed an overestimation in EE in 60-90% of subjects by all seven devices tested during resistance exercise, but were contradictory to the findings by Bai et al. (40), who observed an underestimation in EE measurements by four different activity monitors. However, Bai et al. (40) observed higher inaccuracy in EE measurements by all devices during resistance exercise compared to aerobic exercise and sedentary activities.

There are a number of factors that can affect the accuracy of EE measurements by activity monitors, including the device wear location and the algorithms used to calculate EE. In addition, the ability of the reference standard to accurately measure anaerobic metabolism should be considered when comparing activity monitors to reference standards during exercise.

The location at which an activity monitor is worn during different types of activity can affect the accuracy of EE measurements, with some devices displaying decreased accuracy on upper limbs compared to lower limbs. Some activities, such as cycling and resistance exercise, may require some limbs to be active, while others remain inactive. If activity monitors are being worn on limbs that are primarily inactive, no movement will be detected by the built-in accelerometers, which could result in the activity monitors reporting an underestimation in EE. Although an underestimation in EE was not displayed during resistance exercise by most of the devices tested, the FC2 displayed underestimation during both cycling and resistance exercise. The FC2 is a wrist-worn activity monitor and little to no arm movement was required for cycling and some of the

exercises in our resistance protocol. Studies have observed greater accuracy in EE by research-grade activity monitors worn at the waist compared to those worn at the wrist or ankle during resistance exercise, due to the lack of consistent upper or lower limb movement occurring during resistance exercise (92,93). The lack of arm movement detected by the FC2's built-in accelerometer may have resulted in the underestimation in EE that was observed during cycling and resistance exercise. Although the algorithms being used are proprietary and there is no way of knowing for certain, the FC2 may be relying heavily on accelerometry data to calculate EE, which may not be effective for measuring EE during cycling, resistance, and other exercises that involve more lower limb movement and require little to no upper limb movement.

Since the EE algorithms being used by consumer-grade activity monitors are proprietary, it is difficult to know how EE is being calculated and how EE measurement accuracy can be improved. Despite the algorithms being proprietary, Fitbit and Garmin both use age, height, weight, and sex to estimate EE in all of their devices (94,95). However, incorporating other anthropometrics into EE algorithms may help to further improve EE measurement accuracy by these activity monitors. For example, Plasqui et al. (96) determined that incorporating lean body mass and fat mass into EE measurements from a research-grade activity monitor during physical activity improved the accuracy of EE measurements. Incorporating these and other anthropometrics into EE algorithms may also help to improve the accuracy of consumer-grade activity monitors at measuring EE. Consumer-grade activity monitor manufacturers, such as Fitbit and Garmin, may be using use other inputs to measure EE, but these inputs are undisclosed.

Although the FC2 underestimated EE, the FI, GVHR+, and GV3 all overestimated EE during resistance exercise. Due to the brief, intense bouts of activity performed during resistance exercise, it is considered to be an anaerobic activity in nature. Therefore, using aerobic measurement tools, such as indirect calorimetry, to measure EE during this activity may not be appropriate. Only measuring aerobic metabolism and not accounting for anaerobic metabolism during anaerobic activities will likely result in an underestimation in EE measurements (97). If there is an underestimation in EE measurements by the reference standard (i.e. indirect calorimeter) that the activity monitors are being compared against, it may make it seem as though the activity monitors are significantly overestimating EE, when in fact, their EE measurements may be more accurate than they appear. Scott (97) suggested that measuring excess post-exercise oxygen consumption (EPOC) may be a more effective method for assessing EE for resistance exercise than only measuring oxygen consumption during this activity. Vezina et al. (98) applied this to EE measurements taken by an indirect calorimeter during resistance exercise and compared them to EPOC-adjusted EE measurements. They discovered that EPOC-adjusted EE measurements were significantly higher than non-adjusted EE measurements taken from indirect calorimetry ($P < 0.001$) (98). If the EPOC-adjusted EE measurements are higher than the non-adjusted EE measurements from indirect calorimetry, then the magnitude of error between the EPOC-adjusted EE measurements and the EE measurements from activity monitors will decrease, and the activity monitors may no longer display a significant overestimation in EE measurements during resistance exercise.

An unexpected finding of Study B was the discovery of a significant overestimation in EE measurements by the GV3 in individuals with T1D compared to individuals without T1D. No other study to date has assessed the difference in accuracy of activity monitors between individuals with and without T1D. This observation needs to be examined in more detail prior to incorporating any activity monitor into an AP system since it may make insulin dosing less accurate if EE is not measured accurately. Specifically, less insulin delivery would normally be initiated by an AP system as EE increases, at least to a point (25), since muscle contraction/exercise is known to recruit non-insulin-mediated glucose transport (99) and increase the risk of hypoglycemia in individuals with T1D (25). In other words, if activity monitors are incorporated into AP systems under the assumption that they will provide the same level of accuracy in individuals with T1D as individuals without T1D, and they do not, it may affect the accuracy of the AP system and cause it to make inaccurate decisions. In this case, it is difficult to know exactly why the GV3 overestimated EE in individuals with T1D.

Upon examination, it was discovered that the individuals with T1D in Study B had a significantly higher BF% and a darker skin tone score on the Fitzpatrick skin tone scale than individuals without T1D. As previously discussed, different anthropometrics, such as lean body mass and fat mass, can affect the accuracy of activity monitors at measuring EE (96). Although no relationship was observed between lean body mass and accuracy in EE in this study, there was an overall moderate, positive relationship displayed between %RE in EE for the GV3 and BF% ($r=0.52$, $P<0.0001$), indicating that those with a higher BF% tended to display a greater overestimation in EE by the GV3 compared to those with a lower BF%. If Garmin were to incorporate BF% into their EE

algorithms, it may result in improvements in the EE measurements by the GV3. In addition, studies examining the accuracy of HR measurements by PPG sensors in wrist-worn activity monitors have discovered greater error in HR measurements in individuals with darker skin tones (46,100). Garmin states that their devices that measure HR, such as the GV3, use HR to help calculate EE (95). If the findings from previous studies are also true for the GV3, then the darker skin tones in individuals with T1D may be producing higher error in HR measurements, which could be affecting the accuracy of the EE measurements produced by the GV3 and contributing to the difference in EE accuracy displayed between the two groups of subjects.

Furthermore, PPG sensors used in wrist-worn activity monitors, such as the GV3, measure HR by flashing a light-emitting diode (LED) light through the skin to detect the contraction and dilation of arteries in the wrist with each pulse (101). Studies have demonstrated an increase in endothelial dysfunction exhibited by decreased flow-mediated dilation (widening of arteries in response to increased blood flow) of the brachial artery and the capillaries in fingertips of children, adolescents, and young adults with T1D (102–105). Since PPG sensors are measuring the contraction and dilation of arteries, the impaired vasodilatory response present in individuals with T1D may be affecting the ability of the PPG sensors in the GV3 to accurately measure HR and producing greater error in HR measurements. This could consequently be impacting the accuracy of the EE measurements produced by the GV3 and causing the overestimation in EE displayed by the GV3 in individuals with T1D.

Although no other study has examined the effects of the combination of activities performed in Study B on BG, it was predicted that a greater decrease in BG

concentration would be observed during visit 1 compared to visit 2, based on the response observed to these activities in individuals with T1D in previous studies (58–63). In visit 1, a greater mean decrease of 2.6 ± 3.8 mmol/L in BG was observed after completing three different moderate- to vigorous-intensity exercises, while in visit 2, a mean decrease of 1.3 ± 2.8 mmol/L in BG was observed after completing one low-intensity and one high-intensity activity. A mean increase of 1.0 ± 2.0 mmol/L in BG was observed during the VO_2 peak test in both individuals with and without T1D. This is consistent with the findings of Purdon et al. (63), who also examined an increase in BG in individuals with and without T1D after completing a VO_2 max test, which was attributed to the rise in catecholamines observed during this type of vigorous-intensity activity. The decrease in BG observed during resistance exercise, HIIT-C, and HIIT-T has also been observed in other studies where individuals with T1D performed these types of exercises (58–62). The increased glucose uptake that occurs during exercise, the exogenous insulin present in the body, and the impaired counterregulatory responses in individuals with T1D all contribute to increased glucose utilization by muscles and a decrease in BG during exercise in these individuals (106,107). A small mean decrease in BG (0.1 ± 0.9 mmol/L) was observed during ADL, which was expected since all activities being performed during ADL were at a low-intensity.

6.2. STRENGTHS

There are three main strengths of this project. While most studies only examine the accuracy of activity monitors during steady-state walking and running, this project examined the accuracy of activity monitors during 4-5 different types of activity,

including graded exercise tests, resistance exercise, ADL, and HIIT. Study B was the first study to compare the accuracy of activity monitors between individuals with and without T1D, and as a result, a difference in the accuracy of the GV3 between the two groups of subjects was observed. Study B was also the first study to examine the effects of performing a combination of a VO₂ peak test, resistance exercise, and HIIT-C on BG and of performing ADL and HIIT-T on BG.

6.3. LIMITATIONS

Although there are clear strengths of this project, there were also several limitations. The first limitation is that both Study A (n=20) and Study B (n=15) were underpowered based on our sample size calculations and had a relatively small sample size in comparison to other studies that have assessed the accuracy of consumer-grade activity monitors, with some studies including up to 62 subjects (34). In addition to the small sample size, there was also a lack of diversity in the subjects in both studies. Most subjects were between the ages of 18-29, primarily Caucasian, with a normal BMI. Consequently, this makes the findings of Study A and Study B less generalizable to the rest of the population. Both studies used an in-laboratory exercise protocol, which cannot be generalized to free-living conditions. During the resistance exercise, subjects were allowed to self-select their weights for resistance during each exercise. This can be considered a limitation, because subjects were not required to determine their 1 repetition-maximum (RM) for each exercise and perform exercises at a specific percentage of their 1 RM, meaning the resistance being used was not standardized between subjects. However, by allowing subjects to self-select their weights, it made the

resistance exercise more generalizable to free-living conditions where subjects would self-select their weights based on what they feel comfortable using rather than lifting a certain percentage of their 1 RM. Another limitation of this project is that only four activity monitors were assessed for accuracy in EE, while some other studies have assessed up to nine different activity monitors (39). In addition, there was a lot of data loss due to technical challenges (i.e. device connectedness, etc.) that occurred in both Study A and Study B. Although we had a total of 15-20 subjects complete each study, the data loss that occurred resulted in a much smaller sample of data to analyze for certain activities (e.g. $n=6$ for EE by GVHR+ during VO_2 Peak-T). Furthermore, many manufacturers do not provide access to minute-by-minute data from their devices, so it made it difficult to assess EE data at a higher sample rate for all devices. By not having access to this data, it prevented us from examining how accuracy may have changed throughout the duration of each activity and limited us to only examining total EE accuracy for each activity. By analyzing minute-by-minute data, it would have allowed us to determine if there was greater accuracy or error during certain parts of each activity and provided a better understanding of why there may have been greater accuracy in EE measurements during some activities compared to others.

6.4. FUTURE DIRECTIONS

Consumer-grade activity monitor accuracy is a relatively new area of research and a lot more work still needs to be conducted in this field. The activity monitors assessed in this project, as well as all other activity monitors, need to be tested in a larger, more diverse group of subjects. While most studies have only assessed activity

monitors in laboratory settings, activity monitors should also be examined during exercise in free-living conditions. More studies need to assess activity monitor accuracy during a variety of non-steady-state activities, such as elliptical exercise, cycling, resistance exercise, and different sports. Newer activity monitors are being released to the consumer market every year and these newer monitors also need to be assessed for accuracy (19). For the purpose of this project, only EE accuracy was assessed. However, all metrics measured by activity monitors need to be assessed for accuracy, such as HR, steps, etc. When you are working with any type of technology, there is a high likelihood of data loss. Researchers need to be aware of all of the possible areas of and reasons for data loss to occur and need to make an effort to try and reduce as much data loss as possible during these types of studies. Although manufacturers do not provide consumers with access to minute-by-minute data, this data can be accessed through the third-party, pay for service research platform, Fitabase. Previously, Fitabase only provided access to Fitbit data at a higher sample rate than what was available to consumers from Fitbit, but recently, they have begun providing access to Garmin data as well. Although acquiring a subscription is expensive, if researchers are able to, they should acquire a subscription to Fitabase, so that they can access and analyze activity monitor data at a higher sample rate than what is provided by the manufacturer. This will be especially useful for analyzing HR data from Fitbit and Garmin devices.

Based on the findings from Study A and Study B, there are a number of things that should be examined in future studies. The underestimation in EE displayed by the FC2 during cycling and resistance exercise needs to be examined further. However,

since Study A was conducted, Fitbit has released an updated model of the FC2 to consumers, known as the Fitbit Charge 3 (FC3). Future studies should examine the FC3 to see if the underestimation in EE measurements that was observed during cycling and resistance exercise by the FC2 also exists in the FC3. In addition, researchers should examine if there is greater accuracy on the FC3 when subjects are performing more upper limb movement compared to when they are performing little to no upper limb movement. All previous studies that have examined the accuracy of activity monitors have used indirect calorimetry as the reference standard. However, in future studies, researchers should consider using EPOC-adjusted EE measurements from indirect calorimetry as the reference standard for resistance exercise and any other anaerobic activities to avoid displaying inaccurate overestimations in EE by activity monitors during these types of activities. More research also needs to be conducted using the GV3 in a larger sample of individuals with and without T1D to see if the overestimation in EE displayed by the GV3 in Study B persists in individuals with T1D compared to those without T1D. Furthermore, more research needs to be done to determine if the inaccuracy in EE displayed by the GV3 can be attributed to darker skin tones, higher BF%, and decreased flow-mediated dilation in subjects.

6.5. CONCLUSIONS

The FC2, FI, GVHR+, and GV3 were assessed for accuracy during a variety of physical activities. All four activity monitors displayed a MAPE of $>10\%$, while only the GVHR+ and GV3 displayed a %RE of $\leq 10\%$. The FC2 displayed a significant underestimation in EE measurements during cycling, while the FI, GVHR+, and GV3 all

displayed an overestimation in EE during resistance exercise. Although the GVHR+ and GV3 displayed a %RE of $\leq 10\%$ in measuring EE, the GVHR+ was not assessed in individuals with T1D and the GV3 displayed a significant overestimation in EE in individuals with T1D compared to individuals without T1D. Therefore, based on the findings of this project, none of the devices tested in Study A or Study B can be considered accurate enough or ready to be incorporated into an AP system at this time, although the GV3 is approaching the accuracy (i.e. $\leq 10\%$ %RE) that was initially displayed by earlier CGM systems that are now routinely incorporated into AP prototypes and Health Canada-approved AP systems. More research needs to be conducted with the GV3 in a larger sample of individuals with and without T1D to determine if the overestimation in EE observed in this study exists in a larger group of subjects with T1D, and if it does exist, it needs to be determined why this overestimation is occurring.

7. REFERENCES

1. Chiang JL, Kirkman MS, Laffel LMB, Peters AL. Type 1 diabetes through the life span: A position statement of the American Diabetes Association. *Diabetes Care*. 2014;37:2034–54.
2. Bluestone JA, Herold K, Eisenbarth GS. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature*. 2010;464:1293–1300.
3. Danne T, Nimri R, Hovorka R, Heinemann L, Doyle FJ, Nørgaard K, Kovatchev B, Scharf M, DeVries JH, Hirsch I, et al. International consensus on use of continuous glucose monitoring. *Diabetes Care*. 2017;40:1631–40.
4. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;383:69–82.
5. Rogers MAM, Kim C, Banerjee T, Lee JM. Fluctuations in the incidence of type 1 diabetes in the United States from 2001 to 2015: A longitudinal study. *BMC Med*. 2017;15:199.
6. Juvenile Diabetes Research Foundation (JDRF). Type 1 Diabetes [Internet]. Available from: <https://www.jdrf.ca/who-we-are/type-1-diabetes/>
7. American Diabetes Association. 7. Diabetes Technology: Standards of Medical Care in Diabetes—2019. *Diabetes Care*. 2019;42:S71–S80.
8. Pickup JC. Insulin-pump therapy for type 1 diabetes mellitus. *N Engl J Med*. 2012;366:1616–24.
9. Benjamin EM. Self-monitoring of blood glucose: The basics. *Clin Diabetes*. 2002;20:45–7.
10. Hirsch IB, Armstrong D, Bergenstal RM, Buckingham B, Childs BP, Clarke WL,

- Peters A, Wolpert H. Clinical application of emerging sensor technologies in diabetes management: Consensus guidelines for continuous glucose monitoring (CGM). *Diabetes Technol Ther*. 2008;10:232–46.
11. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes—2019. *Diabetes Care*. 2019;42:S61–70.
 12. Imran SA, Agarwal G, Bajaj HS, Ross S. Targets for glycemic control. *Can J Diabetes*. 2018;42:47–53.
 13. Diabetes Control and Complications Trial Research Group, Nathan D, Genuth S, Lachin J, Cleary P, Crofford O, Davis M, Rand L, Siebert C. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993;329:977–86.
 14. Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) Study Research Group, Gubitosi-Klug RA, Lachin JM, Backlund JYC, Lorenzi GM, Brillon DJ, Orchard TJ. Intensive diabetes treatment and cardiovascular outcomes in type 1 diabetes: The DCCT/EDIC study 30-year follow-up. *Diabetes Care*. 2016;39:686–93.
 15. Sigal RJ, Armstrong MJ, Bacon SL, Boulé NG, Dasgupta K, Kenny GP, Riddell MC. Physical activity and diabetes. *Can J Diabetes*. 2018;42:S54–S63.
 16. Colberg SR, Sigal RJ, Yardley JE, Riddell MC, Dunstan DW, Dempsey PC, Horton ES, Castorino K, Tate DF. Physical activity/exercise and diabetes: A position statement of the American Diabetes Association. *Diabetes Care*. 2016;39:2065–79.

17. McCarthy MM, Funk M, Grey M. Cardiovascular health In adults with type 1 diabetes. *Prev Med (Baltim)*. 2016;91:138–143.
18. Yardley JE, Hay J, Abou-Setta AM, Marks SD, McGavock J. A systematic review and meta-analysis of exercise interventions in adults with type 1 diabetes. *Diabetes Res Clin Pract*. 2014;106:393–400.
19. Henriksen A, Mikalsen MH, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S. Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables. *J Med Internet Res*. 2018;20:e110.
20. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, Hamilton CB, Li LC. Accuracy of Fitbit devices: Systematic review and narrative syntheses of quantitative data. *J Med Internet Res*. 2018;6:e10527.
21. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis. *J Med Internet Res*. 2014;16:e192.
22. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA*. 2015;313:459–60.
23. Aparicio-Ugarriza R, Mielgo-Ayuso J, Benito PJ, Pedrero-Chamizo R, Ara I, González-Gross M. Physical activity assessment in the general population; Instrumental methods and new technologies. *Nutr Hosp*. 2015;31:219–26.
24. Cinar A. Multivariable adaptive artificial pancreas system in type 1 diabetes. *Curr Diab Rep*. 2017;17:88.
25. Riddell MC, Zaharieva DP, Yavelberg L, Cinar A, Jamnik VK. Exercise and the

- development of the artificial pancreas: One of the more difficult series of hurdles. *J Diabetes Sci Technol*. 2015;9:1217–26.
26. Turksoy K, Quinn LT, Littlejohn E, Cinar A. An integrated multivariable artificial pancreas control system. *J Diabetes Sci Technol*. 2014;8:498–507.
 27. Woodman JA, Crouter SE, Bassett DR, Fitzhugh EC, Boyer WR. Accuracy of consumer monitors for estimating energy expenditure and activity type. *Med Sci Sports Exerc*. 2017;49:371–7.
 28. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: Implications for weight management. *PLoS One*. 2016;11:1–11.
 29. Montes J, Young JC, Tandy R, Navalta JW. Fitbit Flex: Energy expenditure and step count evaluation. *J Exerc Physiol*. 2017;20:134–40.
 30. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc*. 2014;46:1840–8.
 31. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. *Br J Sports Med*. 2018;52:844–50.
 32. Hendrikx J, Ruijs LS, Cox LG, Lemmens PM, Schuijers EG, Goris AH. Clinical evaluation of the measurement performance of the Philips Health Watch: A within-person comparative study. *JMIR Mhealth Uhealth*. 2017;5:e10.
 33. Gusmer RJ, Bosch TA, Watkins AN, Ostrem JD, Dengel DR. Comparison of Fitbit Ultra to ActiGraph GT1M for assessment of physical activity in young adults

- during treadmill walking. *Open Sport Med J*. 2014;8:11–5.
34. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR Mhealth Uhealth*. 2017;5:e34.
 35. Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. *J Med Eng Technol*. 2016;40:325–9.
 36. Diaz KM, Krupka DJ, Chang MJ, Shaffer JA, Ma Y, Goldsmith J, Schwartz JE, Davidson KW. Validation of the Fitbit One for physical activity measurement at an upper torso attachment site. *BMC Res Notes*. 2016;9:1–9.
 37. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, Schwartz JE, Davidson KW. Fitbit: An accurate and reliable device for wireless physical activity tracking. *Int J Cardiol*. 2015;185:138–40.
 38. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multisensor physical activity monitors. *PLoS One*. 2017;12:1–15.
 39. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. *Front Physiol*. 2017;8:725.
 40. Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y, Meier NF, Dixon PM. Comparison of consumer and research monitors under semistructured settings. *Med Sci Sports Exerc*. 2016;48:151–8.
 41. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. *J Sports Sci*. 2018;36:1734–41.

42. Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA. Accuracy of the vivofit activity tracker. *J Med Eng Technol.* 2016;40:298–306.
43. Pribyslavska V, Caputo JL, Coons JM, Barry VW. Impact of EPOC adjustment on estimation of energy expenditure using activity monitors. *J Med Eng Technol.* 2018;42:265–73.
44. Thiebaud RS, Funk MD, Patton JC, Massey BL, Shay TE, Schmidt MG, Giovannitti N. Validity of wrist-worn consumer products to measure heart rate and energy expenditure. *Digit Heal.* 2018;4:1–7.
45. Stackpool CM, Porcari JP, Mikat RP, Gillette C, Foster C. The accuracy of various activity trackers in estimating steps taken and energy expenditure. *J Fit Res.* 2014;3:32–48.
46. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, Wheeler MT, Ashley EA. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J Pers Med.* 2017;7:1–12.
47. Sasaki JE, Hickey A, Mavilia M, Tedesco J, John D, Keadle SK, Freedson PS. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. *J Phys Act Heal.* 2015;12:149–54.
48. Price K, Bird SR, Lythgo N, Raj IS, Wong JYL, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *J Med Eng Technol.* 2017;41:208–15.
49. Noah JA, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and

- indirect calorimetry. *J Med Eng Technol*. 2013;37:456–62.
50. Nelson MB, Kaminsky LA, Dickin DC, Montoye AHK. Validity of consumer-based physical activity monitors for specific activity types. *Med Sci Sports Exerc*. 2016;48:1619–28.
 51. Roos L, Taube W, Beeler N, Wyss T. Validity of sports watches when estimating energy expenditure during running. *BMC Sports Sci Med Rehabil*. 2017;9:1–8.
 52. Kendall B, Bellovary B, Gothe NP. Validity of wearable activity monitors for tracking steps and estimating energy expenditure during a graded maximal treadmill test. *J Sports Sci*. 2019;37:42–9.
 53. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, Gillan WW, Gusew EE, Kraemer RR. Validity of wearable activity monitors during cycling and resistance exercise. *Med Sci Sports Exerc*. 2018;50:624–33.
 54. Canadian Society for Exercise Physiology. Canadian Physical Activity Guideline for Adults [Internet]. Available from:
http://www.csep.ca/CMFiles/Guidelines/CSEP_PAGuidelines_adults_en.pdf
 55. American College of Sports Medicine. Resistance Training for Health and Fitness [Internet]. Available from:
<https://www.prescriptiontogetactive.com/app/uploads/resistance-training-ACSM.pdf>
 56. Yavelberg L, Zaharieva D, Cinar A, Riddell MC, Jamnik V. A pilot study validating select research-grade and consumer-based wearables throughout a range of dynamic exercise intensities in persons with and without type 1 diabetes: A novel approach. *J Diabetes Sci Technol*. 2018;12:569–76.

57. Machač S, Procházka M, Radvanský J, Slabý K. Validation of physical activity monitors in individuals with diabetes: Energy expenditure estimation by the multisensor SenseWear Armband Pro3 and the step counter Omron HJ-720 against indirect calorimetry during walking. *Diabetes Technol Ther.* 2013;15:413–8.
58. Yardley JE, Sigal RJ, Kenny GP, Riddell MC, Lovblom LE, Perkins BA. Point accuracy of interstitial continuous glucose monitoring during exercise in type 1 diabetes. *Diabetes Technol Ther.* 2013;15:46–9.
59. Maran A, Pavan P, Bonsembiante B, Brugin E, Ermolao A, Avogaro A, Zaccaria M. Continuous glucose monitoring reveals delayed nocturnal hypoglycemia after intermittent high-intensity exercise in nontrained patients with type 1 diabetes. *Diabetes Technol Ther.* 2010;12:763–8.
60. Iscoe KE, Riddell MC. Continuous moderate-intensity exercise with or without intermittent high-intensity work: Effects on acute and late glycaemia in athletes with type 1 diabetes mellitus. *Diabet Med.* 2011;28:824–32.
61. Guelfi KJ, Jones TW, Fournier PA. Intermittent high-intensity exercise does not increase the risk of early postexercise hypoglycemia in individuals with type 1 diabetes. *Diabetes Care.* 2005;28:416–8.
62. Guelfi KJ, Jones TW, Fournier PA. The decline in blood glucose levels is less with intermittent high-intensity compared with moderate exercise in individuals with type 1 diabetes. *Diabetes Care.* 2005;28:1289–95.
63. Purdon C, Brousson M, Nyveen SL, Miles PD, Halter JB, Vranic M, Marliss EB. The roles of insulin and catecholamines in the glucoregulatory response during

- intense exercise and early recovery in insulin-dependent diabetic and control subjects. *J Clin Endocrinol Metab.* 1993;76:566–73.
64. ePARmed-X+. Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) [Internet]. Available from: <http://eparmedx.com/wp-content/uploads/2013/01/January2018PARQPlus.pdf>
 65. Craig C, Marshall A, Sjöström M, Bauman A, Booth M, Ainsworth B, Pratt M, Ekelund U, Yngve A, Sallis J, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sport Exerc.* 2003;35:1381–95.
 66. National Institutes of Health, National Heart, Lung, and Blood Institute, North American Association for the Study of Obesity. The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. Rockville, MD; 2000.
 67. Malinow K, Lynch J, Thomas S, Friedmann E, Long J. Automated blood pressure recording: The phenomenon of blood pressure elevations during speech. *Angiology.* 1982;33:474–9.
 68. Peters G, Binder S, Campbell N. The effect of crossing legs on blood pressure: A randomized single-blind cross-over study. *Blood Press Monit.* 1999;4:97–101.
 69. Borg G. Psychophysical bases of perceived exertion. *Med Sci Sport Exerc.* 1982;14:377–81.
 70. Gillen JB, Gibala MJ. Is high-intensity interval training a time-efficient exercise strategy to improve health and fitness? *Appl Physiol Nutr Metab.* 2014;39:409–12.
 71. Karvonen MJ, Kentala E, Mustala O. The effects of training on heart rate; A longitudinal study. *Ann Med Exp Biol Fenn.* 1957;35:307–15.

72. Duffield R, Dawson B, Pinnington HC, Wong P. Accuracy and reliability of a Cosmed K4b2 portable gas analysis system. *J Sci Med Sport*. 2004;7:11–22.
73. Schrack JA, Simonsick EM, Ferrucci L. Comparison of the Cosmed K4b2 portable metabolic system in measuring steady-state walking energy expenditure. *PLoS One*. 2010;5:2–6.
74. Perez-Suarez I, Martin-Rincon M, Gonzalez-Henriquez JJ, Fezzardi C, Perez-Regalado S, Galvan-Alvarez V, Juan-Habib JW, Morales-Alamo D, Calbet JAL. Accuracy and precision of the COSMED K5 portable analyser. *Front Physiol*. 2018;9:1–12.
75. Guidetti L, Meucci M, Bolletta F, Emerenziani G, Gallotta M, Baldari C. Validity, reliability and minimum detectable change of COSMED K5 portable gas exchange system in breath-by-breath mode. *PLoS One*. 2018;13:1–12.
76. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, Houghtaling P, Javadikasgari H, Desai MY. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med Sci Sports Exerc*. 2017;49:1697–703.
77. Fitbit, Inc. Fitbit Charge 2 [Internet]. Available from:
<https://www.fitbit.com/eu/shop/charge2?activeFeature=specs>
78. Garmin International, Inc. vivosmart HR+ [Internet]. Available from:
<https://buy.garmin.com/en-CA/CA/p/548743#overview>
79. Jette M, Sidney K, Blumchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation. *Clin Cardiol*. 1990;13:555–65.
80. Swiatoniowski AK, Quillen EE, Shriver MD, Jablonski NG. Technical note: Comparing Von Luschan skin color tiles and modern spectrophotometry for

- measuring human skin pigmentation. *Am J Phys Anthropol.* 2013;151:325–30.
81. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. *Arch Dermatol.* 1988;124:869–71.
 82. Dlugos JF, Taylor JL. Visible reflectance spectroscopy of human skin: The use of CIE L*a*b* color analysis for in vivo ethnic skin characterization. Waltham, MA; 2012.
 83. Fitbit, Inc. Fitbit Ionic [Internet]. Available from: <https://www.fitbit.com/en-ca/ionic>
 84. Garmin International, Inc. vivosmart 3 [Internet]. Available from: <https://buy.garmin.com/en-CA/CA/p/567813#overview>
 85. Hortensius J, Slingerland RJ, Kleefstra N, Logtenberg SJJ, Groenier KH, Houweling ST, Bilo HJG. Self-monitoring of blood glucose: The use of the first or the second drop of blood. *Diabetes Care.* 2011;34:556–60.
 86. Lima J, Mesquita D, Santos K, Fernandes K, Mendonca R, Quinto C, Silva D, Nobrega LH. The impact of not washing hands on the result of capillary glycemia. *J Clin Mol Endocrinol.* 2016;1:1–4.
 87. Greene C, Brown D, Wallace J, Pardo S, Pflug B. Comparative accuracy of 3 blood glucose monitoring systems that communicate with an insulin pump. *Endocr Pract.* 2014;20:1016–21.
 88. Riddell MC, Gallen IW, Smart CE, Taplin CE, Adolfsson P, Lumb AN, Kowalski A, Rabasa-Lhoret R, McCrimmon RJ, Hume C, et al. Exercise management in type 1 diabetes: A consensus statement. *Lancet Diabetes Endocrinol.* 2017;5:377–90.
 89. Dexcom, Inc. Dexcom G5 Mobile Continuous Glucose Monitoring System [Internet]. Available from: <https://www.dexcom.com/en-CA/g5-mobile-canada-en>

90. Freckmann G, Link M, Kamecke U, Haug C, Baumgartner B, Weitgasser R. Performance and usability of three systems for continuous glucose monitoring in direct comparison. *J Diabetes Sci Technol*. 2019;1–9.
91. Welsh JB, Gao P, Derdzinski M, Puhr S, Johnson TK, Walker TC, Graham C. Accuracy, utilization, and effectiveness comparisons of different continuous glucose monitoring systems. *Diabetes Technol Ther*. 2019;21:1–5.
92. Rawson ES, Walsh TM. Estimation of resistance exercise energy expenditure using accelerometry. *Med Sci Sports Exerc*. 2010;42:622–8.
93. Stec MJ, Rawson ES. Estimation of resistance exercise energy expenditure using triaxial accelerometry. *J Strength Cond Res*. 2012;26:1413–22.
94. Fitbit, Inc. How does Fitbit estimate how many calories I've burned? [Internet]. Available from:
https://help.fitbit.com/articles/en_US/Help_article/1381/?q=bmr&l=en_US&fs=Search&pn=1
95. Garmin International, Inc. Calorie Terminology [Internet]. Available from:
<https://support.garmin.com/en-CA/?faq=IkI4cwCLIK7ox362uGQEV7>
96. Plasqui G, Joosen AMCP, Kester AD, Goris AHC, Westerterp KR. Measuring free-living energy expenditure and physical activity with triaxial accelerometry. *Obes Res*. 2005;13:1363–9.
97. Scott CB. Quantifying the immediate recovery energy expenditure of resistance training. *J Strength Cond Res*. 2011;25:1159–63.
98. Vezina JW, Der Ananian CA, Campbell KD, Meckes N, Ainsworth BE. An examination of the differences between two methods of estimating energy

- expenditure in resistance training activities. *J Strength Cond Res.* 2014;28:1026–31.
99. Sylow L, Kleinert M, Richter EA, Jensen TE. Exercise-stimulated glucose uptake - Regulation and implications for glycaemic control. *Nat Rev Endocrinol.* 2017;13:133–48.
100. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. *J Med Eng Technol.* 2015;39:264–71.
101. Maeda Y, Sekine M, Tamura T. The advantages of wearable green reflected photoplethysmography. *J Med Syst.* 2011;35:829–34.
102. Scaramuzza A, Redaelli F, Giani E, Macedoni M, Giudici V, Gazzarri A, Bosetti A, De Angelis L, Zuccotti G. Adolescents and young adults with type 1 diabetes display a high prevalence of endothelial dysfunction. *Acta Paediatr.* 2015;104:192–7.
103. Pillay S, Anderson J, Couper J, Maftei O, Gent R, Peña AS. Children with type 1 diabetes have delayed flow-mediated dilation. *Can J Diabetes.* 2018;42:276–80.
104. Nascimento AMMDAD, Sequeira IJ, Vasconcelos DF, Gandolfi L, Pratesi R, Nóbrega YKDM. Endothelial dysfunction in children with type 1 diabetes mellitus. *Arch Endocrinol Metab.* 2017;61:476–83.
105. Cé GV, Rohde LE, da Silva AMV, Coutinho MKP, de Castro AC, Bertoluci MC. Endothelial dysfunction is related to poor glycemic control in adolescents with type 1 diabetes under 5 years of disease: Evidence of metabolic memory. *J Clin Endocrinol Metab.* 2011;96:1493–9.

106. Cryer PE. Exercise-related hypoglycemia-associated autonomic failure in diabetes. *Diabetes*. 2009;58:1951–2.
107. Ertl AC, Davis SN. Evidence for a vicious cycle of exercise and hypoglycemia in type 1 diabetes mellitus. *Diabetes Metab Res Rev*. 2004;20:124–30.

8. ABBREVIATIONS

ADA	American Diabetes Association
ADL	Activities of daily living
ANOVA	Analysis of variance
AP	Artificial pancreas
API	Application programming interface
BF%	Body fat percentage
BG	Blood glucose
CGM	Continuous glucose monitoring
CSII	Continuous subcutaneous insulin infusion
CVD	Cardiovascular disease
EE	Energy expenditure
EPOC	Excess post-exercise oxygen consumption
FC2	Fitbit Charge 2
FC3	Fitbit Charge 3
FI	Fitbit Ionic
GPS	Global positioning system
GVHR+	Garmin vivosmart HR+
GV3	Garmin vivosmart 3
HbA _{1c}	Glycated hemoglobin
HIIT-C	High-intensity interval training on a cycle ergometer
HIIT-T	High-intensity interval training on a treadmill
HR	Heart rate
HRR	Heart rate reserve
HSD	Honestly significant difference
ICC	Intraclass correlation coefficient
IPAQ	International Physical Activity Questionnaire
LED	Light-emitting diode
LoA	Limits of agreement
MAPE	Mean absolute percentage error
MDI	Multiple daily injections
METs	Metabolic equivalents
PAR-Q+	Physical Activity Readiness Questionnaire for Everyone
PPG	Photoplethysmography
RM	Repetition-maximum
RPE	Rating of perceived exertion
SMBG	Self-monitoring of blood glucose
T1D	Type 1 diabetes
T1DEXI	Type 1 Diabetes and Exercise Initiative
T2D	Type 2 diabetes
VCO ₂	Volume of carbon dioxide production
VO ₂	Volume of oxygen consumption
VO ₂ max	Maximum volume of oxygen consumption
VO ₂ Peak	Peak volume of oxygen consumption
VO ₂ Peak-C	Progressive aerobic exercise test on a cycle ergometer

VO₂ Peak-T
WC
%RE

Progressive aerobic exercise test on a treadmill
Waist circumference
Mean relative percentage error

9. TABLES & FIGURES

9.1. SUMMARY OF THE SCIENTIFIC LITERATURE

Table 1. Summary of Literature on Accuracy of Activity Monitors for Walking. Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for walking.

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Alsubheen et al. (2016) (42)	Walking at 0% incline	Garmin vivofit	Wrist	-45%					
	Walking at 5% incline			-30%					
	Walking at 10% incline			-18%					
Chowdhury et al. (2017) (38)	Walking at 2.5 or 3 mph	Apple Watch	Wrist		24%				
		Fitbit Charge HR	Wrist		73%				
		Jawbone UP24	Wrist		27%				
		Microsoft Band	Wrist		37%				
	Walking at 2.5 or 3 mph with weighted bags (6kg for females, 10kg for males)	Apple Watch	Wrist		14%				
		Fitbit Charge HR	Wrist		48%				
		Jawbone UP24	Wrist		14%				
		Microsoft Band	Wrist		49%				
Diaz et al. (2015) (37)	Walking at 3 mph	Fitbit Flex	Wrist	52%					
	Walking at 4 mph			33%					
Diaz et al. (2016) (36)	Walking at 1.9 mph	Fitbit Flex	Wrist	83%					
		Fitbit One	Hip	-8%					
		Fitbit One	Upper Torso	-10%					
	Walking at 3 mph	Fitbit Flex	Wrist	68%					
		Fitbit One	Hip	13%					
		Fitbit One	Upper Torso	18%					
	Walking at 4 mph	Fitbit Flex	Wrist	29%					
		Fitbit One	Hip	12%					
		Fitbit One	Upper Torso	20%					

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Dondzila & Garner (2016) (35)	Walking at 3 mph	Fitbit Charge	Wrist	21%					
		Jabra Sport Pulse Earbuds	Ear	-27%					
	Walking at 4 mph	Fitbit Charge	Wrist	-11%				0.52	
		Jabra Sport Pulse Earbuds	Ear	-10%				0.52	
Dooley et al. (2017) (34)	Walking at 2.5 mph	Apple Watch	Wrist		20%				
		Fitbit Charge HR	Wrist		85%				
		Garmin Forerunner 225	Wrist		78%				
	Walking at 3.5 mph	Apple Watch	Wrist		14%				
		Fitbit Charge HR	Wrist		46%				
		Garmin Forerunner 225	Wrist		74%				
Gusmer et al. (2014) (33)	Slow walking (avg. speed: 2.5 mph)	Fitbit Ultra	Hip					0.69	
	Brisk walking (avg. speed: 3.0 mph)							0.94	
Nelson et al. (2016) (50)	Self-paced walking (avg. speed: 2.9 mph)	Fitbit Flex	Wrist		53%				
		Fitbit One	Hip		31%				
		Fitbit Zip	Hip		68%				
		Jawbone UP24	Wrist		24%				
Noah et al. (2013) (49)	Walking (3.5 mph, 0% incline)	Fitbit	Hip						0.70
		Fitbit Ultra	Hip						0.83
	Walking (3.5 mph, 5% incline)	Fitbit	Hip						0.72
		Fitbit Ultra	Hip						0.81

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Pribyslavská (2018) (43)	Walking at moderate-intensity (40-59% HRR)	Fitbit Surge	Wrist	-4%	14%				
		Garmin vívofit	Wrist	-12%	20%				
Price et al. (2017) (48)	Walking at 1.6-4.0 mph and 1% incline	Fitbit One	Hip			2.2 kcal/min	-2.3 to 6.8 kcal/min	0.73	
		Garmin vívofit	Wrist			-1.7 kcal/min	-4.1 to 0.7 kcal/min	0.71	
		Jawbone UP	Wrist			-2.1 kcal	-21.9 to 17.7 kcal	0.72	
Sasaki et al. (2015) (47)	Walking at 3 mph	Fitbit	Hip	-22%					
	Walking at 4 mph			- ~35%					
	Walking with weighted bags (4 lb each)			~40%					
Stackpool et al. (2014) (45)	Self-paced walking (3.0-4.2 mph)	Adidas miCoach	Not reported	34%				0.53	
		BodyMedia FIT CORE	Not reported					0.68	
		Fitbit Ultra	Not reported					0.24	
		Jawbone UP	Not reported					0.87	
		Nike FuelBand	Not reported					0.49	
Thiebaud et al. (2018) (44)	Walking at 2 mph	Fitbit Surge	Wrist	45%		2.7 kcal	4.9 to 12.5 kcal	0.52	
		Microsoft Band	Wrist	-9%		-0.7 kcal	2.0 to 8.8 kcal	0.19	
		TomTom Cardio	Wrist	0.4%		0.1 kcal	2.7 to 9.6 kcal	0.57	
	Walking at 3 mph	Fitbit Surge	Wrist	62%		10.3 kcal	19.2 to 36.4 kcal	0.53	
		Microsoft Band	Wrist	-4%		-0.9 kcal	7.8 to 24.7 kcal	0.24	
		TomTom Cardio	Wrist	12%		2.2 kcal	11.9 to 26.6 kcal	0.70	
	Walking at 4 mph	Fitbit Surge	Wrist	53%		17.0 kcal	38.6 to 63.4 kcal	0.55	
		Microsoft Band	Wrist	-2%		-1.3 kcal		0.15	
		TomTom Cardio	Wrist	3%		1.0 kcal	17.8 to 47.4 kcal	0.66	

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Wahl et al. (2017) (39)	Walking at 2.7 mph	Beurer AS80	Wrist	17%			-12.9 to 18.1 kcal		-0.17
		Fitbit Charge	Wrist	75%			7.4 to 26.8 kcal		0.15
		Fitbit Charge HR	Wrist	83%			4.2 to 35.0 kcal		0.17
		Garmin Forerunner 920XT	Wrist	-27%			-21.1 to 8.0 kcal		0.35
		Garmin vívoactive	Wrist	4%			-5.0 to 6.4 kcal		0.85
		Garmin vívofit	Wrist	1%			-8.2 to 8.1 kcal		0.73
		Garmin vivosmart	Wrist	3%			-5.0 to 6.5 kcal		0.91
		Polar Loop	Wrist	56%			-6.4 to 29.5 kcal		-0.11
		Withings Pulse Ox	Hip	-10%			-14.6 to 7.8 kcal		0.44
		Withings Pulse Ox	Wrist	-11%			-13.9 to 6.6 kcal		0.28
	Walking at 4.3 mph	Beurer AS80	Wrist	-19%			-33.5 to 12.0 kcal		-0.09
		Fitbit Charge	Wrist	34%			-8.2 to 37.2 kcal		0.29
		Fitbit Charge HR	Wrist	18%			-7.3 to 22.9 kcal		0.58
		Garmin Forerunner 920XT	Wrist	-17%			-33.4 to 15.5 kcal		0.25
		Garmin vívoactive	Wrist	37%			-15.3 to 48.3 kcal		0.28
		Garmin vívofit	Wrist	19%			-20.2 to 35.5 kcal		0.35
		Garmin vivosmart	Wrist	36%			-20.4 to 51.6 kcal		0.21
		Polar Loop	Wrist	54%			-23.1 to 67.6 kcal		0.02
		Withings Pulse Ox	Hip	19%			-31.5 to 48.1 kcal		-0.15
		Withings Pulse Ox	Wrist	17%			-36.8 to 44.5 kcal		-0.15

Table 2. Summary of Literature on Accuracy of Activity Monitors for Running. Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for running.

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Chowdhury et al. (2017) (38)	Running at 5.2 mph	Apple Watch	Wrist		22%				
		Fitbit Charge HR	Wrist		16%				
		Jawbone UP24	Wrist		29%				
		Microsoft Band	Wrist		18%				
Diaz et al. (2016) (36)	Running at 5.2 mph	Fitbit Flex	Wrist	25%					
		Fitbit One	Hip	-3%					
		Fitbit One	Upper Torso	11%					
Dondzila & Garner (2016) (35)	Running at 5 mph	Fitbit Charge	Wrist	-14%					
		Jabra Sport Pulse Earbuds	Ear	-16%				0.51	
	Running at 6 mph	Fitbit Charge	Wrist	-23%					
		Jabra Sport Pulse Earbuds	Ear	-23%					
Dooley et al. (2017) (34)	Running at 5.5 mph	Apple Watch	Wrist		17%				
		Fitbit Charge HR	Wrist		20%				
		Garmin Forerunner 225	Wrist		31%				
Kendall et al. (2019) (52)	Graded maximal test	Basis B1	Wrist						0.80
		Fitbit Flex	Wrist						0.81
		Jawbone UP	Hip						0.75
		OMRON HJ-321	Hip						0.83
		Polar FT7	Chest						0.74
Nelson et al. (2016) (50)	Self-paced running (avg. speed: 4.5 mph)	Fitbit Flex	Wrist		35%				
		Fitbit One	Hip		22%				
		Fitbit Zip	Hip		37%				
		Jawbone UP24	Wrist		46%				

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Noah et al. (2013) (49)	Running at 5.5 mph	Fitbit	Hip					0.56	
		Fitbit Ultra	Hip					0.87	
Pribyslavská (2018) (43)	Running at vigorous-intensity (60-84% HRR)	Fitbit Surge	Wrist	-13%					
		Garmin vívofit	Wrist	-12%	18%				
Price et al. (2017) (48)	Running at 5.0-7.5 mph and 1% incline	Fitbit One	Hip			3.6 kcal/min	-7.8 to 14.9 kcal/min		
		Garmin vívofit	Wrist			-1.5 kcal/min	-7.7 to 4.8 kcal/min	0.35	
		Jawbone UP	Wrist			20.6 kcal	-37.7 to 79.0 kcal	0.65	
Roos et al. (2017) (51)	Running at 30% VO ₂ peak	Garmin Forerunner 920XT	Wrist		26%				
		Polar V800	Wrist		23%				
		Suunto Ambit2	Wrist		42%			0.30	
	Running at 50% VO ₂ peak	Garmin Forerunner 920XT	Wrist		18%				
		Polar V800	Wrist		11%				
		Suunto Ambit2	Wrist		36%				
	Running at 70% VO ₂ peak	Garmin Forerunner 920XT	Wrist		12%				
		Polar V800	Wrist		10%				
		Suunto Ambit2	Wrist		21%				
	Running at 90% VO ₂ peak	Garmin Forerunner 920XT	Wrist		38%				
		Polar V800	Wrist		30%				
		Suunto Ambit2	Wrist		23%				
	Running at 110% VO ₂ peak	Garmin Forerunner 920XT	Wrist		49%				
		Polar V800	Wrist		40%				
		Suunto Ambit2	Wrist		32%				
Sasaki et al. (2015) (47)	Running at 5.5 mph	Fitbit	Hip	- ~5%					

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Stackpool et al. (2014) (45)	Self-paced running (5.0-8.5 mph)	Adidas miCoach	Not reported					0.81	
		BodyMedia FIT CORE	Not reported	-13%				0.73	
		Fitbit Ultra	Not reported					0.63	
		Jawbone UP	Not reported	20%				0.69	
		Nike FuelBand	Not reported	15%				0.71	
Thiebaud et al. (2018) (44)	Running at 5 mph	Fitbit Surge	Wrist	37%		20.7 kcal	60.8 to 99.2 kcal	0.53	
		Microsoft Band	Wrist	4%		1.1 kcal	34.6 to 84.7 kcal	0.20	
		TomTom Cardio	Wrist	-16%		-9.5 kcal	29.7 to 68.4 kcal	0.60	
	Running at 6 mph	Fitbit Surge	Wrist	25%		21.7 kcal	86.6 to 138.7 kcal	0.60	
		Microsoft Band	Wrist	8%		5.4 kcal	61.9 to 130.3 kcal	0.37	
		TomTom Cardio	Wrist	-27%		-24.0 kcal	40.5 to 92.9 kcal	0.56	
Wahl et al. (2017) (39)	Running at 6.3 mph	Beurer AS80	Wrist	-33%			-50.7 to 6.5 kcal		-0.04
		Fitbit Charge	Wrist	14%			-8.2 to 21.7 kcal		0.68
		Fitbit Charge HR	Wrist	20%			-16.4 to 38.4 kcal		0.43
		Garmin Forerunner 920XT	Wrist	-9%			-39.8 to 25.1 kcal		0.25
		Garmin vivoactive	Wrist	20%			0.8 to 22.8 kcal		0.69
		Garmin vívofit	Wrist	7%			-14.1 to 20.0 kcal		0.76
		Garmin vívosmart	Wrist	24%			-19.1 to 48.0 kcal		0.46
		Polar Loop	Wrist	51%			-17.3 to 68.3 kcal		-0.09
		Withings Pulse Ox	Hip	9%			-33.0 to 39.1 kcal		0.27
		Withings Pulse Ox	Wrist	5%			-32.1 to 34.3 kcal		0.20

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation	ICC
Wahl et al. (2017) (39)	Running at 8.1 mph	Beurer AS80	Wrist	-43%			-68.6 to 1.4 kcal		-0.02
		Fitbit Charge	Wrist	-0.1%			-22.3 to 18.8 kcal		0.73
		Fitbit Charge HR	Wrist	22%			-22.1 to 51.5 kcal		0.43
		Garmin Forerunner 920XT	Wrist	-11%			-50.8 to 30.5 kcal		0.28
		Garmin vivoactive	Wrist	6%			-15.6 to 23.0 kcal		0.82
		Garmin vívofit	Wrist	-11%			-35.9 to 17.3 kcal		0.56
		Garmin vívosmart	Wrist	9%			-30.0 to 42.2 kcal		0.59
		Polar Loop	Wrist	41%			-30.3 to 76.3 kcal		-0.25
		Withings Pulse Ox	Hip	-8%			-43.5 to 32.0 kcal		0.54
		Withings Pulse Ox	Wrist	-5%			-37.3 to 23.4 kcal		0.46
	Outdoor running at 6.3 mph	Beurer AS80	Wrist	-48%			-216.9 to 1.3 kcal		-0.04
		Fitbit Charge	Wrist	-5%			-75.6 to 46.2 kcal		0.64
		Fitbit Charge HR	Wrist	-12%			-99.5 to 40.2 kcal		0.53
		Garmin Forerunner 920XT	Wrist	-21%			-124.5 to 29.3 kcal		0.34
		Garmin vivoactive	Wrist	-5%			-46.2 to 24.3 kcal		0.91
		Garmin vívofit	Wrist	-20%			-86.6 to -1.9 kcal		0.56
		Garmin vívosmart	Wrist	-2%			-66.8 to 59.0 kcal		0.82
		Polar Loop	Wrist	22%			-94.8 to 163.0 kcal		-0.18
		Withings Pulse Ox	Hip	-6%			-132.2 to 97.3 kcal		0.21
		Withings Pulse Ox	Wrist	-5%			-130.4 to 91.7 kcal		0.22

Table 3. Summary of Literature on Accuracy of Activity Monitors for Elliptical Exercise. Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for elliptical exercise.

Authors	Testing Protocol	Device Tested	Wear Location	%RE	Pearson Correlation
Stackpool et al. (2014) (45)	Self-paced exercise on elliptical cross-trainer	BodyMedia FIT CORE	Not reported	-20%	0.47
		Fitbit Ultra	Not reported		0.41
		Jawbone UP	Not reported		0.40
		Nike FuelBand	Not reported	-27%	0.08

Table 4. Summary of Literature on Accuracy of Activity Monitors for Cycling. Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for cycling.

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Median Absolute % Error	ICC
Boudreaux et al. (2018) (53)	Graded exercise test on cycle ergometer	Apple Watch 2	Wrist		21%		0.23
		Fitbit Blaze	Wrist		72%		0.20
		Fitbit Charge 2	Wrist		75%		0.18
		Garmin vivosmart HR	Wrist		63%		0.41
		Polar A360	Chest		38%		0.28
		Polar H7	Wrist		30%		0.27
		TomTom Touch	Wrist		41%		0.30
Chowdhury et al. (2017) (38)	Steady-state cycling on cycle ergometer (females at 75 W, males at 100 W)	Apple Watch	Wrist		20%		
		Fitbit Charge HR	Wrist		53%		
		Jawbone UP24	Wrist		82%		
		Microsoft Band	Wrist		38%		
Nelson et al. (2016) (50)	Self-paced cycling on cycle ergometer	Fitbit Flex	Wrist		44%		
		Fitbit One	Hip		43%		
		Fitbit Zip	Hip		46%		
		Jawbone UP24	Wrist		57%		

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Median Absolute % Error	ICC
Pribyslavskaya (2018) (43)	Steady-state cycling at moderate-intensity (40-59% HRR)	Fitbit Surge	Wrist	-20%	28%		
		Garmin vívofit	Wrist	-6%	19%		
	Steady-state cycling at vigorous-intensity (60-84% HRR)	Fitbit Surge	Wrist	-23%	28%		
		Garmin vívofit	Wrist	0%	16%		
Sasaki et al. (2015) (47)	Steady-state cycling at 50 W on cycle ergometer	Fitbit	Hip	-104%			
Shcherbina et al. (2017) (46)	Steady-state cycling on cycle ergometer at low-intensity (50-100 W) and high-intensity (80-225 W)	Apple Watch	Wrist			~40%	
		Basis Peak	Wrist			~60%	
		Fitbit Surge	Wrist			~25%	
		Microsoft Band	Wrist			~35%	
		PulseOn	Wrist			~100%	

Table 5. Summary of Literature on Accuracy of Activity Monitors for Resistance Exercise.

Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for resistance exercise.

Authors	Testing Protocol	Device Tested	Wear Location	MAPE	Mean Difference	ICC
Bai et al. (2016) (40)	Self-selected weights, sets, and reps	Fitbit Flex	Wrist	32%	25.5 kcal	
		Jawbone UP24	Wrist	53%	47.2 kcal	
		Misfit Shine	Wrist		30.8 kcal	
		Nike+ FuelBand SE	Wrist		26.3 kcal	
Boudreaux et al. (2018) (53)	3 circuits of 10 reps of leg curls, chest press, leg extensions, and lat pulldowns	Apple Watch 2	Wrist	43%		0.18
		Fitbit Blaze	Wrist	49%		0.14
		Fitbit Charge 2	Wrist	48%		0.08
		Garmin vivosmart HR	Wrist	57%		0.18
		Polar A360	Chest	53%		0.13
		Polar H7	Wrist	43%		0.11
		TomTom Touch	Wrist	52%		0.02

Table 6. Summary of Literature on Accuracy of Activity Monitors for Sports and Sports-Related Exercise. Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for sports and sports-related exercise.

Authors	Testing Protocol	Device Tested	Wear Location	%RE	Mean Difference	95% LoA	Pearson Correlation	ICC
Sasaki et al. (2015) (47)	Basketball, golf, and tennis	Fitbit	Hip		-2.1 kcal	-26.0 to 22.0 kcal		
Stackpool et al. (2014) (45)	Basketball-related activities	Adidas miCoach	Not reported	-60%			0.57	
		BodyMedia FIT CORE	Not reported	-18%			0.56	
		Fitbit Ultra	Not reported	-17%			0.67	
		Jawbone UP	Not reported	-30%			0.65	
		Nike FuelBand	Not reported	-14%			0.47	
Wahl et al. (2017) (39)	Intermittent running simulating a soccer game	Beurer AS80	Wrist	-46%		-82.8 to -8.8 kcal		0.00
		Fitbit Charge	Wrist	2%		-28.5 to 30.0 kcal		0.58
		Fitbit Charge HR	Wrist	26%		-27.7 to 78.1 kcal		0.43
		Garmin Forerunner 920XT	Wrist	-9%		-67.8 to 46.9 kcal		0.22
		Garmin vívoactive	Wrist	-1%		-31.5 to 29.4 kcal		0.74
		Garmin vívofit	Wrist	-21%		-53.5 to 11.4 kcal		0.54
		Garmin vívosmart	Wrist	2%		-61.6 to 67.0 kcal		0.43
		Polar Loop	Wrist	6%		-42.5 to 46.5 kcal		-0.30
		Withings Pulse Ox	Hip	-49%		-72.1 to 3.7 kcal		0.01
		Withings Pulse Ox	Wrist	-39%		-75.5 to -14.0 kcal		0.11

Table 7. Summary of Literature on Accuracy of Activity Monitors for Activities of Daily Living.

Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for activities of daily living.

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	95% LoA	Pearson Correlation
Bai et al. (2018) (41)	Folding laundry, sweeping, moving light boxes, stretching, and slow walking	Apple Watch	Wrist	4%	23%			0.66
		Fitbit Charge HR	Wrist	-60%	61%			0.41
Chowdhury et al. (2017) (38)	Loading and unloading dishwasher, sweeping, and ascending and descending stairs	Apple Watch	Wrist		29-34%			
		Fitbit Charge HR	Wrist		15-39%			
		Jawbone UP24	Wrist		29-52%			
		Microsoft Band	Wrist		45-62%			
Nelson et al. (2016) (50)	Standing, dusting, sweeping, vacuuming, folding laundry, making bed, picking up items from floor, and gardening	Fitbit Flex	Wrist		21%			
		Fitbit One	Hip		27%			
		Fitbit Zip	Hip		36%			
		Jawbone UP24	Wrist		35%			
Sasaki et al. (2015) (47)	Dusting, gardening, laundry, raking, and vacuuming	Fitbit	Hip			-3.1 kcal	-11.0 to 5.2 kcal	

Table 8. Summary of Literature on Accuracy of Activity Monitors for Sedentary Activities.

Summary of the existing scientific literature on accuracy of consumer-grade activity monitors at measuring energy expenditure compared to indirect calorimetry for sedentary activities.

Authors	Testing Protocol	Device Tested	Wear Location	%RE	MAPE	Mean Difference	Pearson Correlation
Bai et al. (2016) (40)	Reading, cell phone use, typing, watching videos, and listening to music/radio	Fitbit Flex	Wrist			5.5 kcal	
		Jawbone UP24	Wrist			11.0 kcal	
		Misfit Shine	Wrist		18%	5.7 kcal	
		Nike+ FuelBand SE	Wrist		20%	6.5 kcal	
Bai et al. (2018) (41)	Computer use, reading, and cell phone use	Apple Watch	Wrist	9%	22%		0.45
		Fitbit Charge HR	Wrist	21%	35%		0.40
Nelson et al. (2016) (50)	Lying down, computer use, watching TV, writing, playing cards, and reading	Fitbit Flex	Wrist		14%		
		Fitbit One	Hip		13%		
		Fitbit Zip	Hip		16%		
		Jawbone UP24	Wrist	-8%	17%		

Table 9. Summary of Literature on Accuracy of Activity Monitors in Individuals with Diabetes.

Summary of the existing scientific literature on accuracy of consumer-grade and research-grade activity monitors at measuring energy expenditure compared to indirect calorimetry in individuals with type 1 diabetes.

Authors	Device Tested	Wear Location	Testing Protocol	%RE	Mean Difference	95% LoA	Pearson Correlation
Machac et al. (2013) (57)	OMRON HJ-720	Hip	Walking at 1.9 mph and 0% incline	71%			0.77
	SenseWear Armband Pro3	Arm		81%			0.79
	OMRON HJ-720	Hip	Walking at 2.5 mph and 0% incline	76%			0.52
	SenseWear Armband Pro3	Arm		78%			0.63
	OMRON HJ-720	Hip	Walking at 3.1 mph and 5% incline	-7%			0.75
	SenseWear Armband Pro3	Arm		-8%			0.74
Yavelberg et al. (2018) (56)	Metria IH1	Arm	Walking at $\leq 50\%$ VO_2 max	4%	0.3 kcal/min	-2.9 to 3.5 kcal/min	
			Moderate-to-vigorous intensity (50-75% VO_2 max) exercise circuit	6%	0.6 kcal/min	-2.6 to 3.9 kcal/min	
			Vigorous-to-maximal intensity (75-100% VO_2 max) exercise circuit	-13%	-1.8 kcal/min	-7.2 to 3.7 kcal/min	

9.2. STUDY A

Table 10. Participant Characteristics. Participant characteristics for Study A (n=20).

Characteristic	Value
M / F	9 / 11
Age (yr)	27.5 \pm 6.0
Height (cm)	173.2 \pm 9.5
Weight (kg)	67.9 \pm 10.8
Body Mass Index (kg/m ²)	22.5 \pm 2.3
VO ₂ Peak (mL/min/kg)	48.0 \pm 8.7
Wrist (cm)	15.6 \pm 2.0
Race, n (%)	
Caucasian	17 (85)
Asian	2 (10)
Native American/Canadian	1 (5)

Table 11. Energy Expenditure. Pooled energy expenditure (EE) data for the different types of activities performed during Study A. Data are shown for each activity type for the Fitbit Charge 2 and Garmin vivosmart HR+. Sample size, mean \pm SD EE reported by each device, mean \pm SD EE reported by the reference standard, mean \pm SD difference between the device measurement and the reference standard, mean relative difference \pm SD (%) between the device measurement and the reference standard, mean absolute difference \pm SD (%) between the device measurement and the reference standard, and correlation between the device measurement and the reference standard are presented in the table.

Measures	Fitbit Charge 2	Garmin vivosmart HR+
VO₂ Peak Test (Treadmill)		
N	10	6
Device Mean \pm SD	192.1 \pm 47.2	216.5 \pm 55.3
Criterion Mean \pm SD	237.3 \pm 72.5	260.5 \pm 77.2
Mean Difference \pm SD	-45.2 \pm 44.4	-44.0 \pm 90.1
Mean Relative % Error \pm SD	-17.0 \pm 14.6	-11.4 \pm 33.7
Mean Absolute % Error \pm SD	19.4 \pm 11.0	28.8 \pm 17.2
Pearson's Correlation	0.81	0.11
VO₂ Peak Test (Cycle Ergometer)		
N	10	9
Device Mean \pm SD	133.6 \pm 77.6	207.0 \pm 48.7
Criterion Mean \pm SD	225.3 \pm 74.7	231.4 \pm 76.5
Mean Difference \pm SD	-91.7 \pm 87.2	-24.4 \pm 63.9
Mean Relative % Error \pm SD	-39.1 \pm 30.6	-4.5 \pm 25.3
Mean Absolute % Error \pm SD	43.5 \pm 23.0	18.9 \pm 16.2
Pearson's Correlation	0.35	0.56
Resistance Exercise		
N	20	16
Device Mean \pm SD	130.2 \pm 46.2	179.8 \pm 56.8
Criterion Mean \pm SD	153.1 \pm 45.5	155.2 \pm 47.8
Mean Difference \pm SD	-22.9 \pm 44.0	24.6 \pm 56.6
Mean Relative % Error \pm SD	-12.9 \pm 29.7	21.0 \pm 35.7
Mean Absolute % Error \pm SD	27.7 \pm 15.9	35.7 \pm 19.7
Pearson's Correlation	0.54	0.43
Activities of Daily Living		
N	20	18
Device Mean \pm SD	103.5 \pm 38.2	100.6 \pm 23.4
Criterion Mean \pm SD	114.4 \pm 25.7	114.8 \pm 27.0
Mean Difference \pm SD	-10.9 \pm 39.4	-14.3 \pm 28.2
Mean Relative % Error \pm SD	-8.8 \pm 29.2	-10.6 \pm 19.3
Mean Absolute % Error \pm SD	20.9 \pm 21.8	17.0 \pm 13.7
Pearson's Correlation	0.29	0.38
High-Intensity Interval Training (Treadmill)		
N	10	9
Device Mean \pm SD	211.1 \pm 57.0	226.9 \pm 58.1
Criterion Mean \pm SD	246.6 \pm 71.9	249.7 \pm 75.6
Mean Difference \pm SD	-35.5 \pm 34.6	-22.8 \pm 61.7
Mean Relative % Error \pm SD	-13.1 \pm 12.7	-4.7 \pm 29.3
Mean Absolute % Error \pm SD	14.5 \pm 10.9	25.0 \pm 13.4
Pearson's Correlation	0.88	0.60

High-Intensity Interval Training (Cycle Ergometer)		
N	10	9
Device Mean \pm SD	128.2 \pm 60.4	205.8 \pm 76.4
Criterion Mean \pm SD	232.8 \pm 44.2	234.9 \pm 46.4
Mean Difference \pm SD	-104.6 \pm 83.8	-29.1 \pm 80.2
Mean Relative % Error \pm SD	-41.9 \pm 31.3	-11.2 \pm 30.8
Mean Absolute % Error \pm SD	41.9 \pm 31.3	26.7 \pm 17.0
Pearson's Correlation	-0.26	0.22

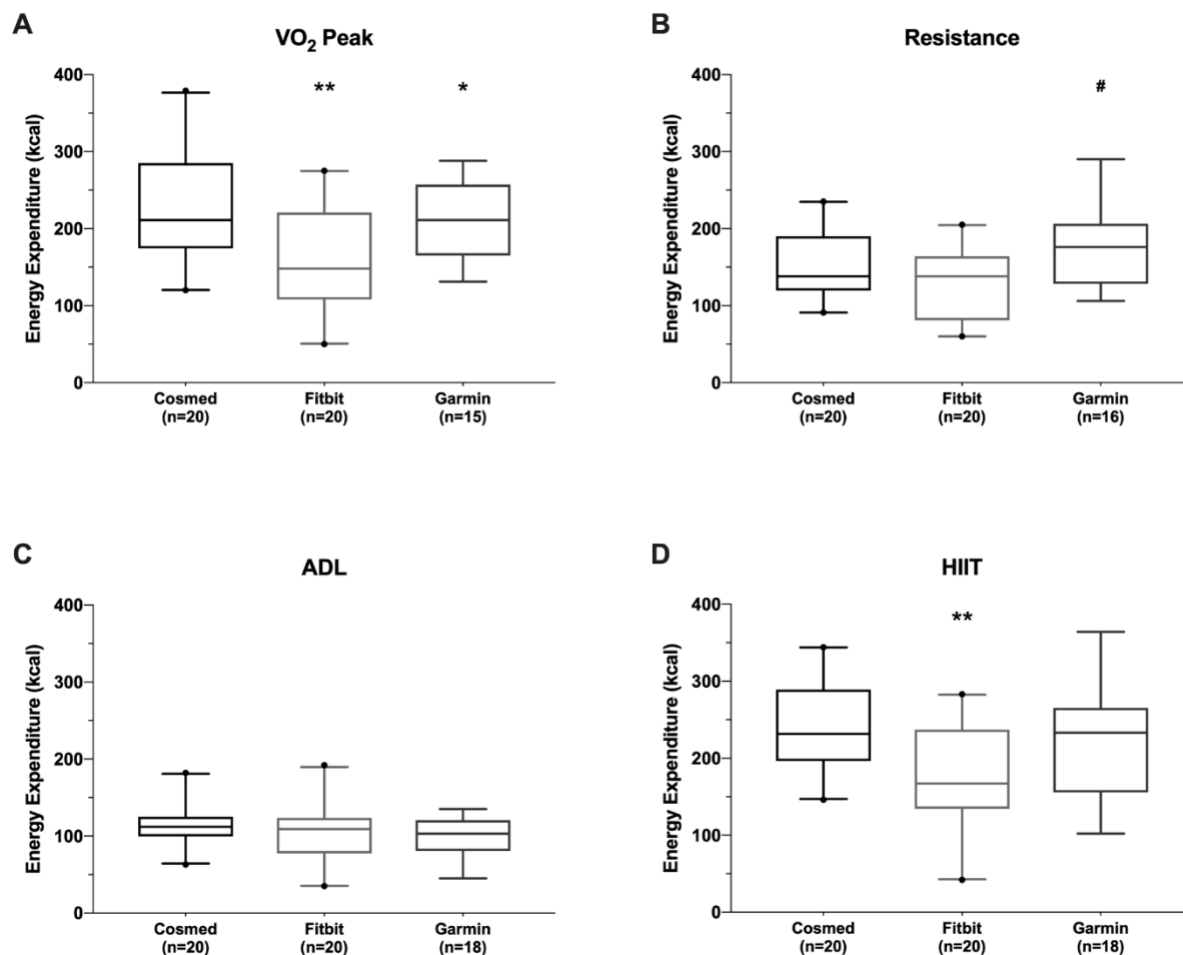


Figure 1. Total Energy Expenditure. Total energy expenditure (kcal) measurements by the Cosmed K4b²/K5, Fitbit Charge 2, and Garmin vivosmart HR+ during the VO₂ peak test (A), resistance exercise (B), activities of daily living (ADL) (C), and high-intensity interval training (HIIT) (D) in Study A. Each box-whisker plot consists of a box that extends from the 25% to the 75% quantile, with a line in the middle of the box representing the median (50% quantile). Each box has error bars that extend to the 5% and 95% quantiles, with outliers displayed with closed circles. ***P*<0.01 compared to Cosmed. **P*<0.05 compared to Fitbit. #*P*<0.0001 compared to Fitbit.

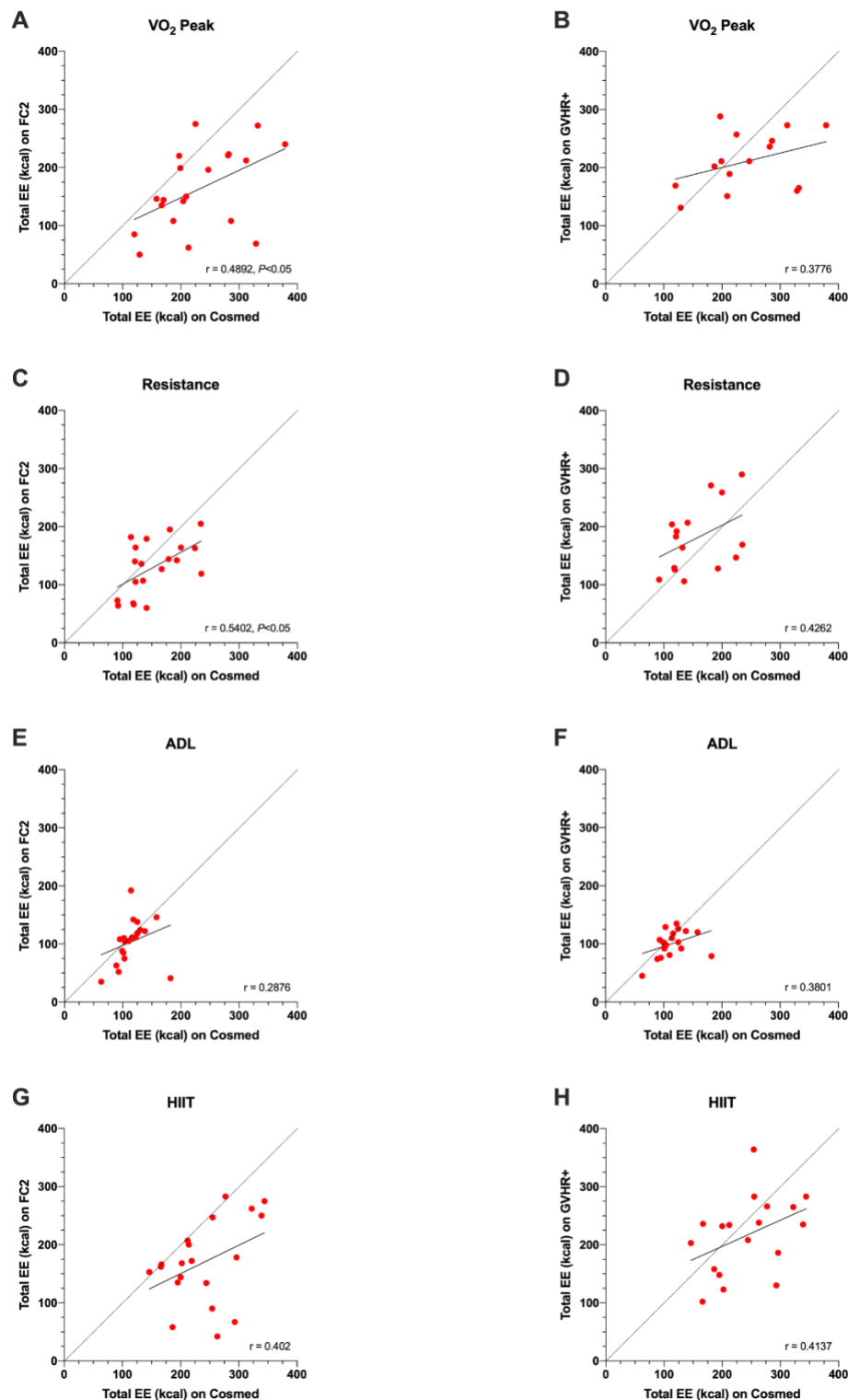


Figure 2. Correlation Plots for Total Energy Expenditure. Correlation plots comparing total energy expenditure (kcal) measurements from the Fitbit Charge 2 (FC2) (A,C,E,G) and Garmin vivosmart HR+ (GVHR+) (B,D,F,H) relative to the Cosmed K4b²/K5 during the VO₂ peak test (A-B), resistance exercise (C-D), activities of daily living (ADL) (E-F), and high-intensity interval training (HIIT) (G-H) in Study A. The black solid line represents linear regression. The grey dotted line indicates the line of identity. n=6-20.

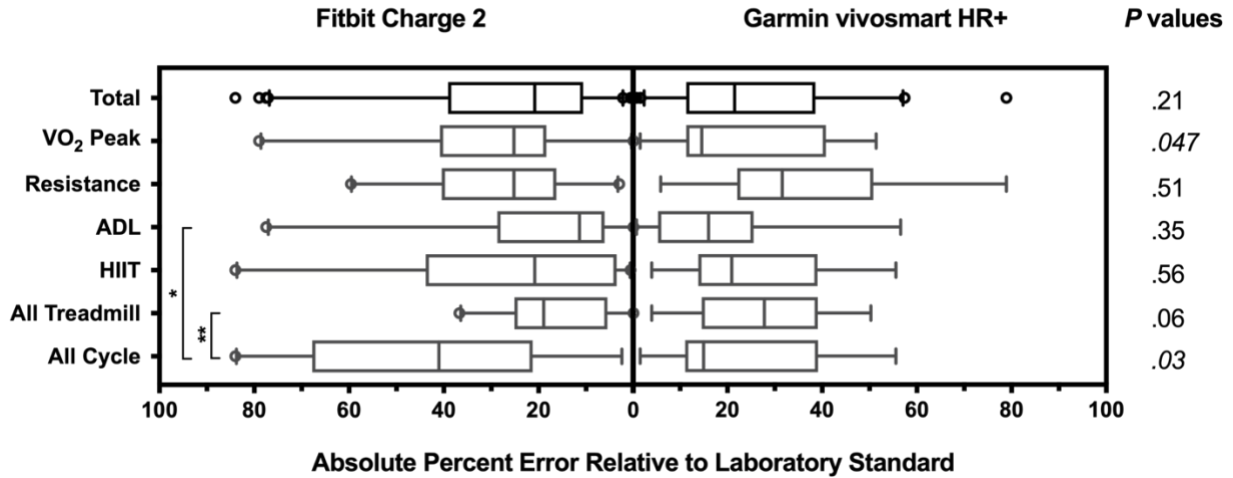


Figure 3. Absolute Percent Error in Energy Expenditure. Absolute percent error of the Fitbit Charge 2 (FC2) and Garmin vivosmart HR+ (GVHR+) in measuring energy expenditure (EE) (kcal) during the VO₂ peak test, resistance exercise, activities of daily living (ADL), and high-intensity interval training (HIIT) in Study A. Each box-whisker plot consists of a box that extends from the 25% to the 75% quantile, with a line in the middle of the box representing the median (50% quantile). Each box has error bars that extend to the 5% and 95% quantiles, with outliers displayed with open circles. The *P* values listed on the right side display the difference in absolute percent error for EE between the FC2 and GVHR+ during each activity. *n*=15-20. **P*<0.05. ***P*<0.01.

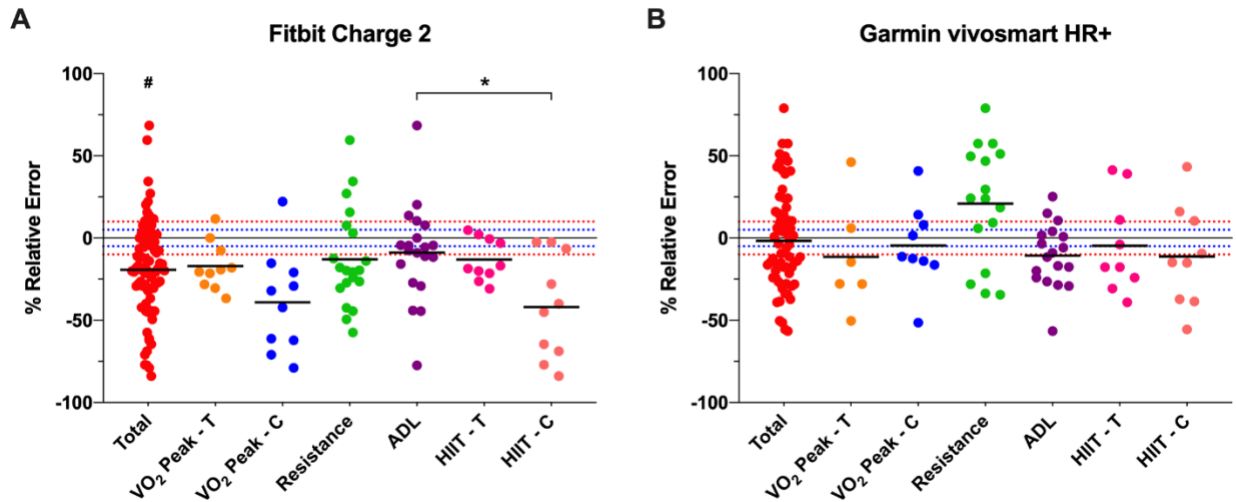


Figure 4. Relative Percent Error in Energy Expenditure. Relative percent error of the Fitbit Charge 2 (A) and Garmin vivosmart HR+ (B) in measuring energy expenditure (kcal) during the VO₂ peak test on a treadmill (VO₂ Peak-T) and cycle ergometer (VO₂ Peak-C), resistance exercise, activities of daily living (ADL), and high-intensity interval training on a treadmill (HIIT-T) and cycle ergometer (HIIT-C) in Study A. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. *n*=6-20. **P*<0.05. #*P*<0.0001 compared to Garmin.

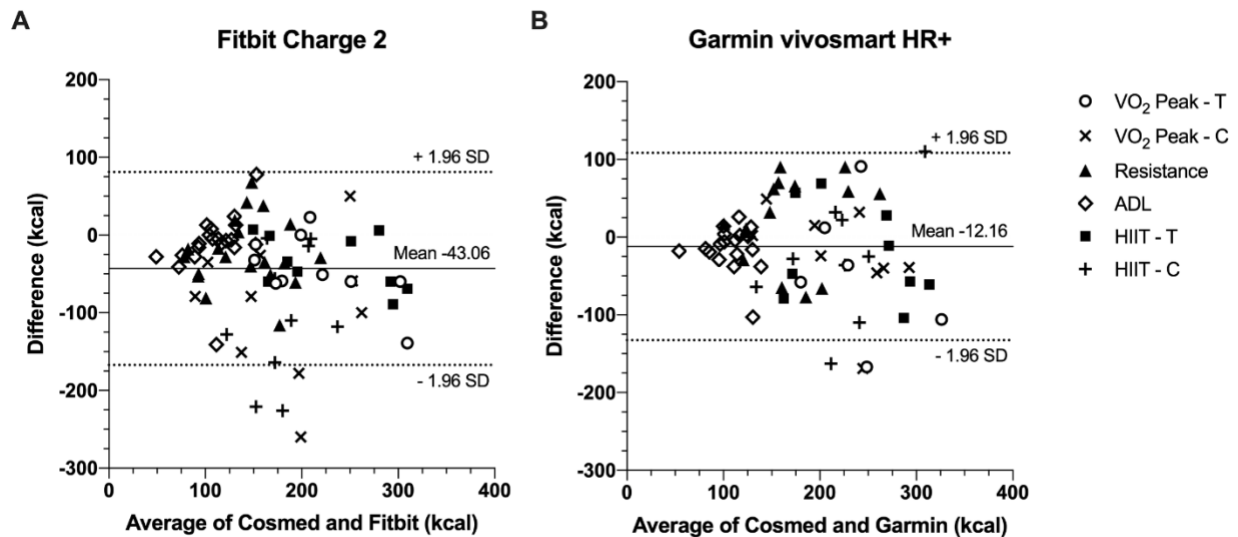


Figure 5. Bland-Altman Plots for Energy Expenditure. Bland-Altman plots comparing energy expenditure (EE) (kcal) measurements from the Fitbit Charge 2 (FC2) (A) and Garmin vivosmart HR+ (GVHR+) (B) relative to the Cosmed K4b²/K5 for all data in Study A, with each type of exercise represented by a different symbol. Mean EE of the Cosmed and FC2 or GVHR+ is displayed on the x-axis and the difference between the Cosmed and FC2 or GVHR+ is displayed on the y-axis. The black solid line indicates the mean difference (bias) between measurements and the black dashed lines indicate the limits of agreement. n=6-20.

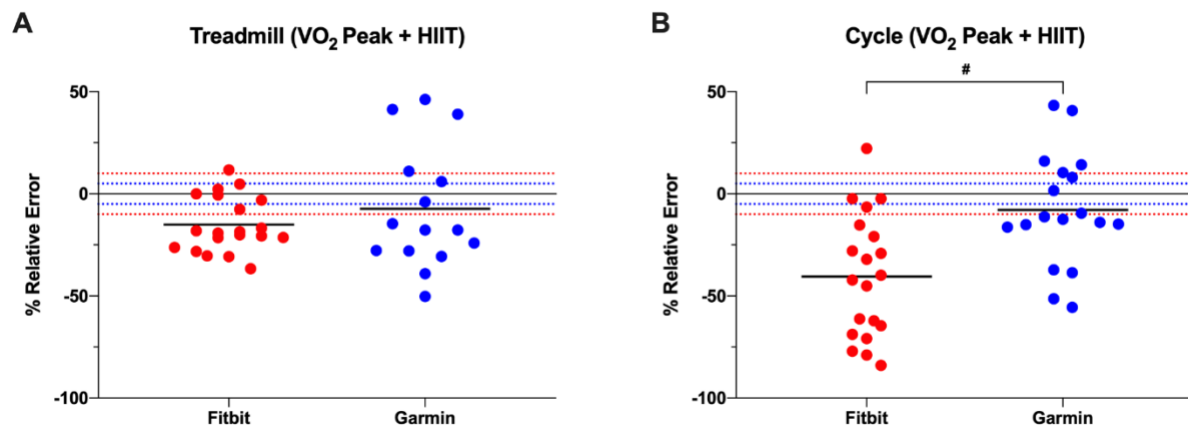


Figure 6. Relative Percent Error in Energy Expenditure for Treadmill and Cycle Ergometer Exercise. Relative percent error of the Fitbit Charge 2 and Garmin vivosmart HR+ in measuring energy expenditure (kcal) during the VO₂ peak test and high-intensity interval training (HIIT) on the treadmill (A) and cycle ergometer (B) in Study A. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. n=9-20. #*P*<0.0001.

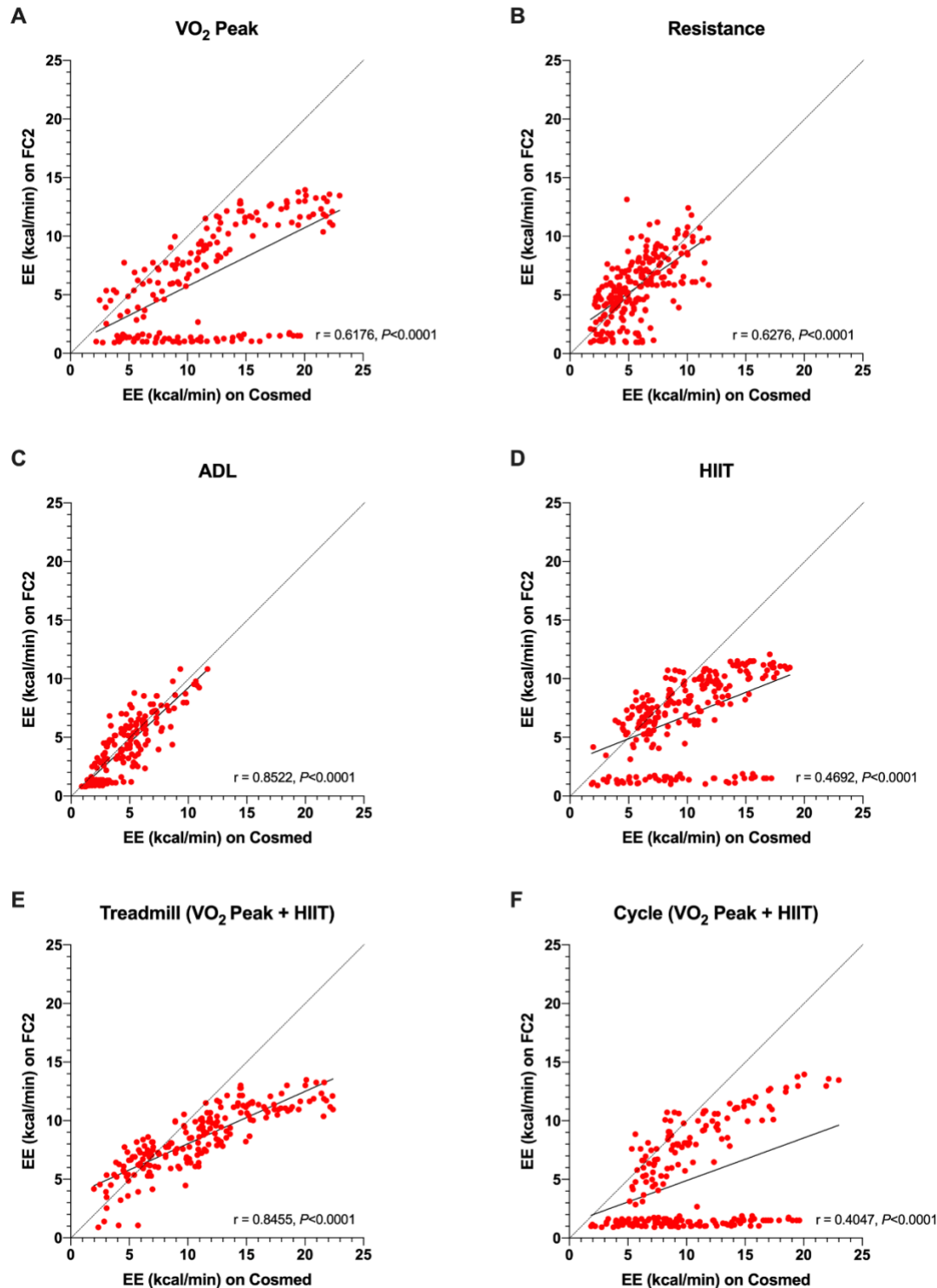


Figure 7. Correlation Plots for Energy Expenditure. Correlation plots comparing energy expenditure (kcal/min) measurements from the Fitbit Charge 2 (FC2) relative to the Cosmed K4b²/K5 during the VO₂ peak test (A), resistance exercise (B), activities of daily living (ADL) (C), and high-intensity interval training (HIIT) (D), as well as all treadmill exercises combined (E) and all cycle ergometer exercises combined (F), in Study A. The black solid line identifies linear regression. The grey dotted line indicates the line of identity. n=10.

9.3. STUDY B

Table 12. Participant Characteristics. Participant characteristics for Study B (n=15).

Characteristic	Total (n=15)	T1D (n=9)	Control (n=6)
M / F	6 / 9	2 / 7	4 / 2
Age (yr)	25.8 ± 8.1	24.9 ± 10.4	27.2 ± 2.1
Height (cm)	170.1 ± 11.1	168.9 ± 11.1	172.0 ± 11.9
Weight (kg)	70.0 ± 13.5	71.8 ± 13.0	67.3 ± 15.1
Body Mass Index (kg/m ²)	24.1 ± 3.4	25.0 ± 3.0	22.6 ± 3.6
VO ₂ Peak (mL/min/kg)	42.9 ± 8.2	41.3 ± 5.9	45.3 ± 10.9
Wrist (cm)	15.8 ± 1.4	16.0 ± 1.6	15.7 ± 1.3
Race, n (%)			
Caucasian	14 (93)	9 (100)	5 (83)
African American	1 (7)	0 (0)	1 (17)

Table 13. Energy Expenditure. Pooled energy expenditure (EE) data for the different types of activities performed during Study B. Data are shown for each activity type for the Fitbit Ionic and Garmin vivosmart 3. Sample size, mean \pm SD EE reported by each device, mean \pm SD EE reported by the reference standard, mean \pm SD difference between the device measurement and the reference standard, mean relative difference \pm SD (%) between the device measurement and the reference standard, mean absolute difference \pm SD (%) between the device measurement and the reference standard, and correlation between the device measurement and the reference standard are presented in the table.

Measures	Fitbit Ionic	Garmin vivosmart 3
VO₂ Peak Test (Treadmill)		
N	12	12
Device Mean \pm SD	251.4 \pm 110.7	249.9 \pm 111.3
Criterion Mean \pm SD	251.5 \pm 71.8	221.6 \pm 64.3
Mean Difference \pm SD	0.1 \pm 53.5	-28.3 \pm 61.9
Mean Relative % Error \pm SD	7.4 \pm 22.6	-4.6 \pm 19.9
Mean Absolute % Error \pm SD	18.6 \pm 14.0	16.4 \pm 11.2
Pearson's Correlation	0.92	0.89
Resistance Exercise		
N	13	11
Device Mean \pm SD	164.2 \pm 47.0	168.2 \pm 42.5
Criterion Mean \pm SD	254.5 \pm 99.3	205.6 \pm 73.2
Mean Difference \pm SD	90.3 \pm 74.1	37.5 \pm 83.2
Mean Relative % Error \pm SD	56.1 \pm 44.4	28.6 \pm 49.7
Mean Absolute % Error \pm SD	56.1 \pm 44.4	44.6 \pm 34.3
Pearson's Correlation	0.70	0.04
Activities of Daily Living		
N	13	13
Device Mean \pm SD	112.8 \pm 23.4	112.8 \pm 23.4
Criterion Mean \pm SD	129.2 \pm 44.6	123.8 \pm 51.7
Mean Difference \pm SD	16.4 \pm 28.4	11.0 \pm 42.5
Mean Relative % Error \pm SD	12.9 \pm 24.2	9.7 \pm 35.0
Mean Absolute % Error \pm SD	21.6 \pm 16.1	23.3 \pm 27.1
Pearson's Correlation	0.83	0.59
High-Intensity Interval Training (Treadmill)		
N	12	13
Device Mean \pm SD	259.0 \pm 53.4	257.8 \pm 51.3
Criterion Mean \pm SD	238.3 \pm 54.6	248.6 \pm 74.2
Mean Difference \pm SD	-20.8 \pm 29.6	-9.2 \pm 69.9
Mean Relative % Error \pm SD	-7.8 \pm 10.5	-2.2 \pm 26.1
Mean Absolute % Error \pm SD	11.0 \pm 6.6	20.1 \pm 15.8
Pearson's Correlation	0.85	0.43
High-Intensity Interval Training (Cycle Ergometer)		
N	12	10
Device Mean \pm SD	230.3 \pm 70.8	240.9 \pm 72.6
Criterion Mean \pm SD	237.7 \pm 77.8	242.4 \pm 79.4
Mean Difference \pm SD	7.4 \pm 68.2	1.5 \pm 95.3
Mean Relative % Error \pm SD	6.8 \pm 32.9	5.8 \pm 43.1
Mean Absolute % Error \pm SD	28.5 \pm 15.6	34.3 \pm 24.2
Pearson's Correlation	0.58	0.22

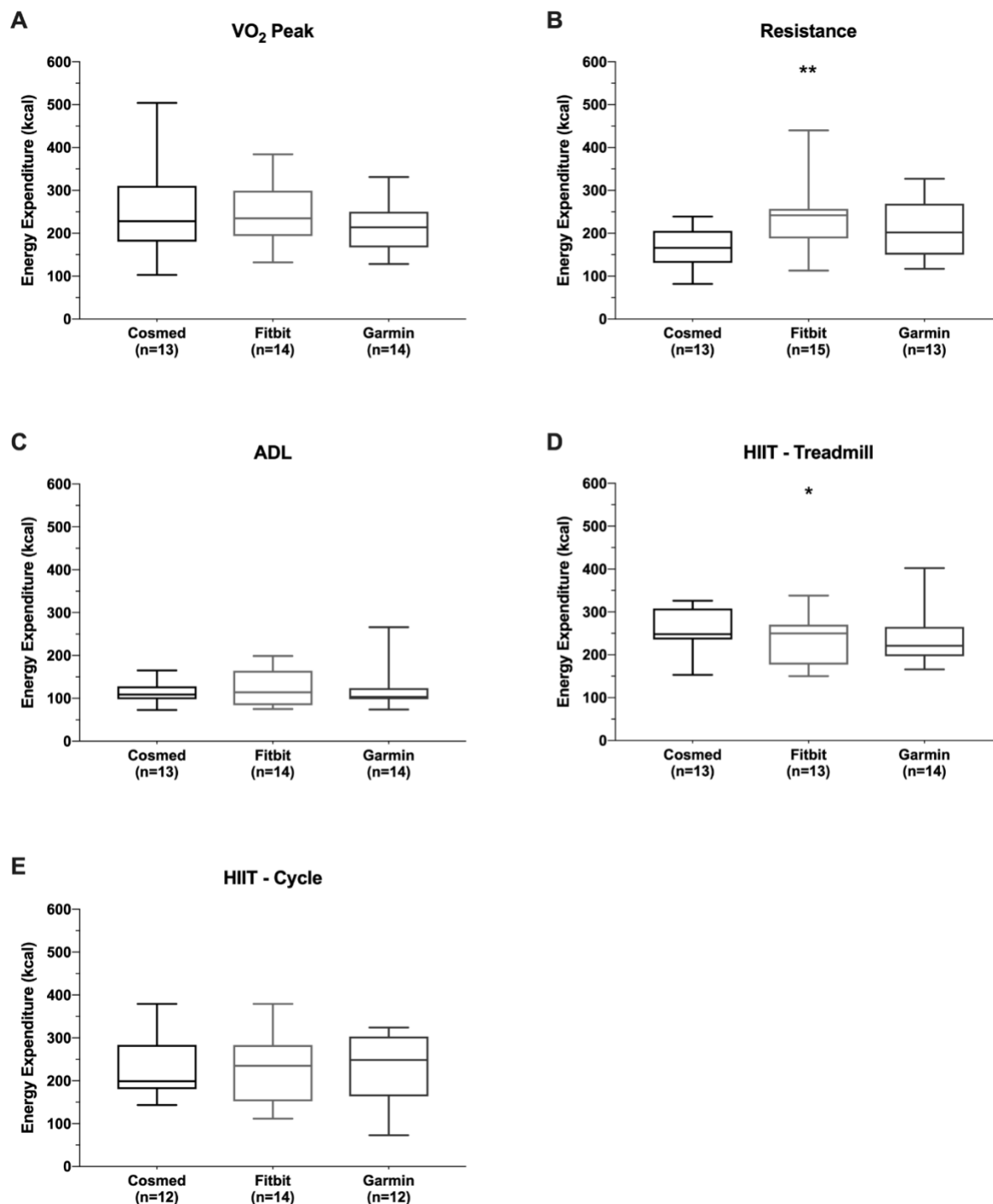


Figure 8. Total Energy Expenditure. Total energy expenditure (kcal) measurements by the Cosmed K5, Fitbit Ionic, and Garmin vivosmart 3 during the VO₂ peak test (A), resistance exercise (B), activities of daily living (ADL) (C), and high-intensity interval training (HIIT) on a treadmill (D) and cycle ergometer (E) in Study B. Each box-whisker plot consists of a box that extends from the 25% to the 75% quantile, with a line in the middle of the box representing the median (50% quantile). Each box has error bars that extend to the 5% and 95% quantiles, with outliers displayed with closed circles. * $P < 0.05$ compared to Cosmed. ** $P < 0.01$ compared to Cosmed.

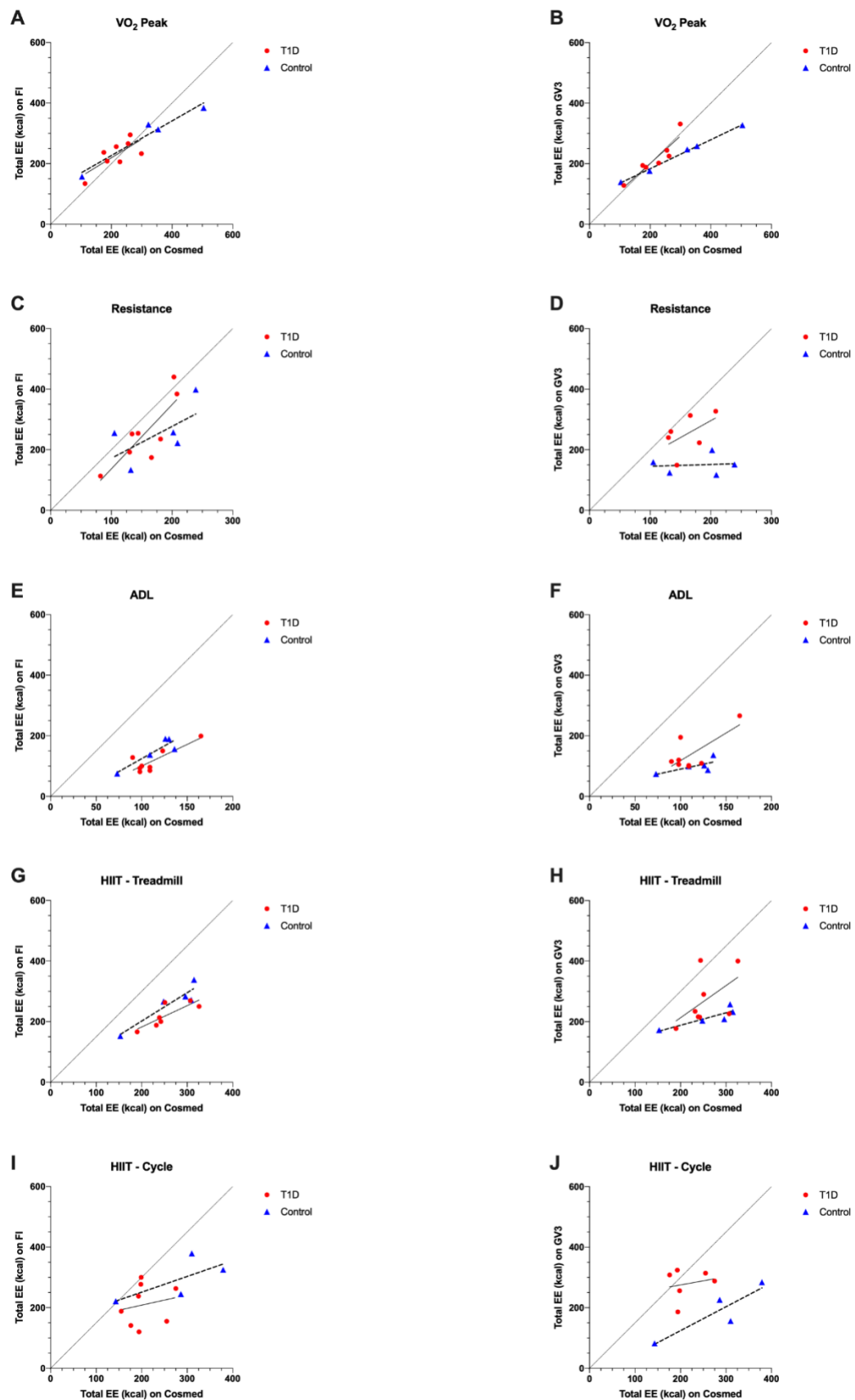


Figure 9. Correlation Plots for Total Energy Expenditure. Correlation plots comparing total energy expenditure (kcal) measurements from the Fitbit Ionic (FI) (A,C,E,G,I) and Garmin vivosmart 3 (GV3) (B,D,F,H,J) relative to the Cosmed K5 during the VO_2 peak test (A-B), resistance exercise (C-D), activities of daily living (ADL) (E-F), and high-intensity interval training (HIIT) on a treadmill (G-H) and cycle ergometer (I-J) in Study B. The black solid line identifies linear regression for T1D subject data and the black dashed line identifies linear regression for control subject data. The grey dotted line indicates the line of identity. n=10-13.

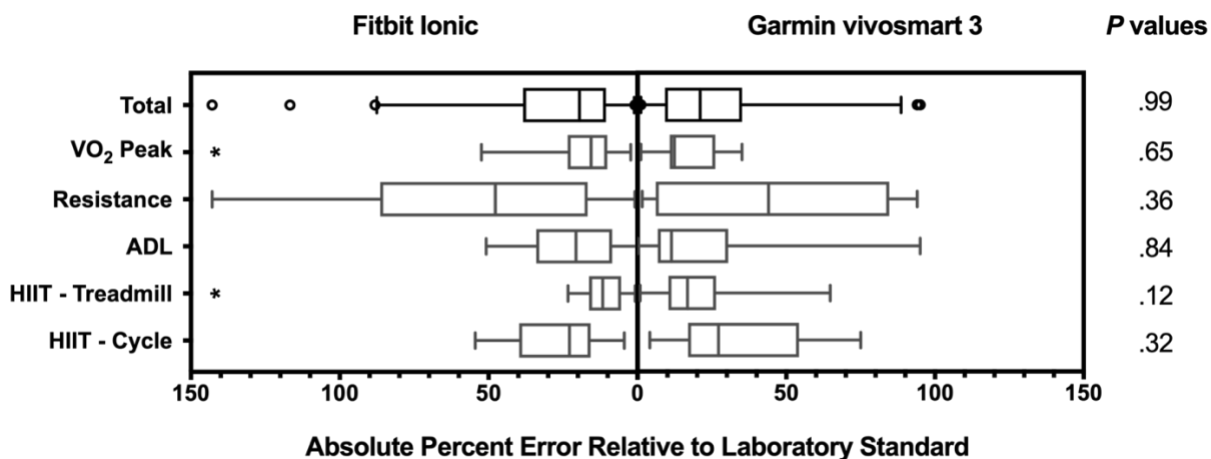


Figure 10. Absolute Percent Error in Energy Expenditure. Absolute percent error of the Fitbit Ionic (FI) and Garmin vivosmart 3 (GV3) in measuring energy expenditure (EE) (kcal) during the VO₂ peak test, resistance exercise, activities of daily living (ADL), and high-intensity interval training (HIIT) on a treadmill and cycle ergometer in Study B. Each box-whisker plot consists of a box that extends from the 25% to the 75% quantile, with a line in the middle of the box representing the median (50% quantile). Each box has error bars that extend to the 5% and 95% quantiles, with outliers displayed with open circles. The *P* values listed on the right side display the difference in absolute percent error for EE between the FI and GV3 during each activity. *n*=10-13. **P*<0.05 compared to Resistance.

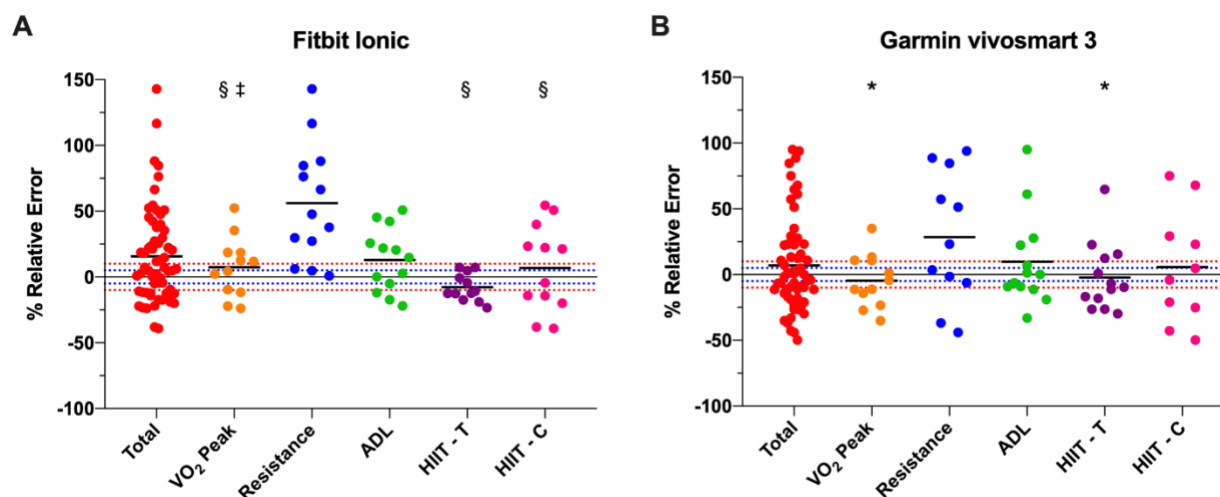


Figure 11. Relative Percent Error in Energy Expenditure. Relative percent error of the Fitbit Ionic (A) and Garmin vivosmart 3 (B) in measuring energy expenditure (kcal) during the VO₂ peak test, resistance exercise, activities of daily living (ADL), and high-intensity interval training on a treadmill (HIIT-T) and cycle ergometer (HIIT-C) in Study B. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. *n*=10-13. ‡*P*<0.05 compared to Garmin. **P*<0.05 compared to Resistance. §*P*<0.001 compared to Resistance.

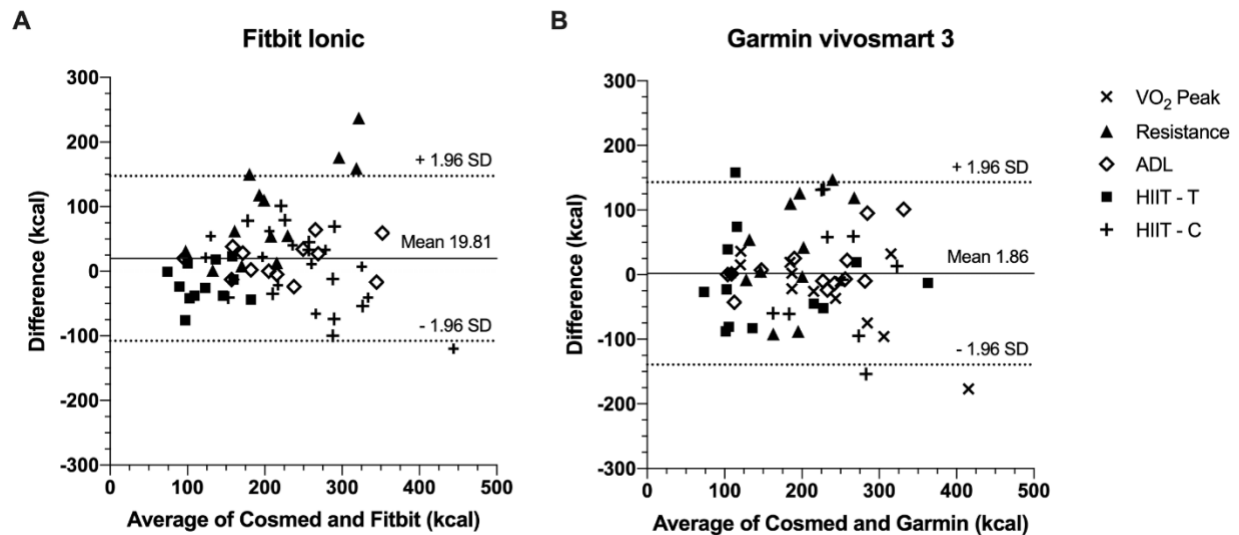


Figure 12. Bland-Altman Plots for Energy Expenditure. Bland-Altman plots comparing energy expenditure (EE) (kcal) measurements from the Fitbit Ionic (FI) (A) and Garmin vivosmart 3 (GV3) (B) relative to the Cosmed K5 for all data in Study B, with each type of exercise represented by a different symbol. Mean EE of the Cosmed and FI or GV3 is displayed on the x-axis and the difference between the Cosmed and FI or GV3 is displayed on the y-axis. The black solid line indicates the mean difference (bias) between measurements and the black dashed lines indicate the limits of agreement. n=10-13.

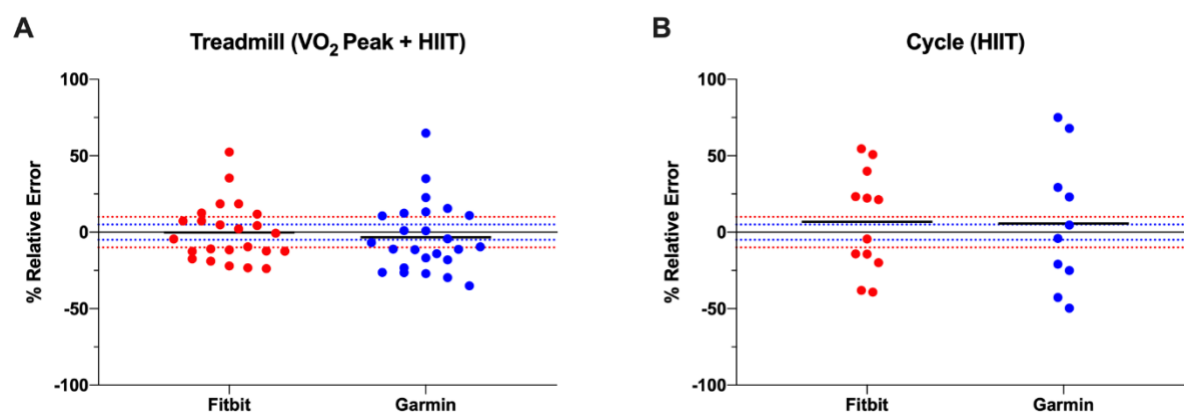


Figure 13. Relative Percent Error in Energy Expenditure for Treadmill and Cycle Ergometer Exercise. Relative percent error of the Fitbit Ionic and Garmin vivosmart 3 in measuring energy expenditure (kcal) during the VO_2 peak test and high-intensity interval training (HIIT) on the treadmill (A) and cycle ergometer (B) in Study B. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. n=10-13.

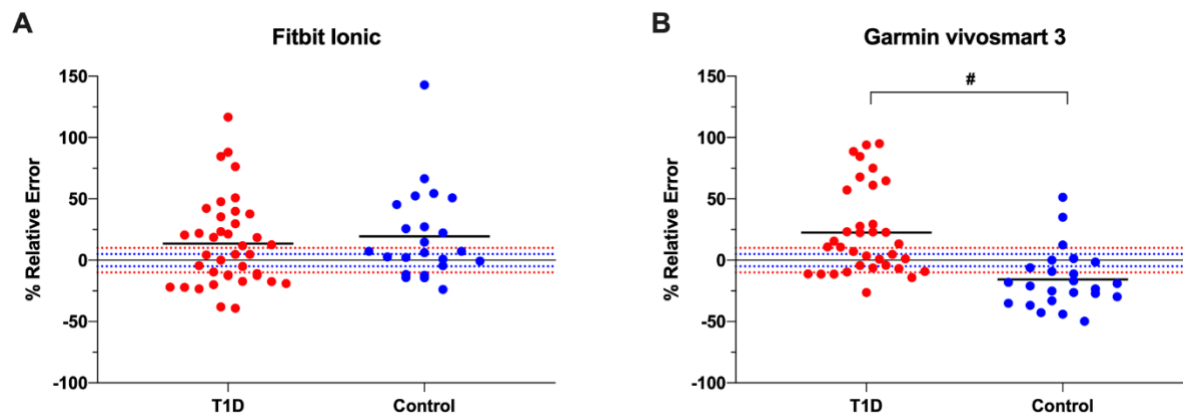


Figure 14. Relative Percent Error in Energy Expenditure in T1D and Control Subjects. Relative percent error of the Fitbit Ionic (A) and Garmin vivosmart 3 (B) in measuring energy expenditure (kcal) in T1D and control subjects in Study B. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. $n=5-9$. # $P<0.0001$.

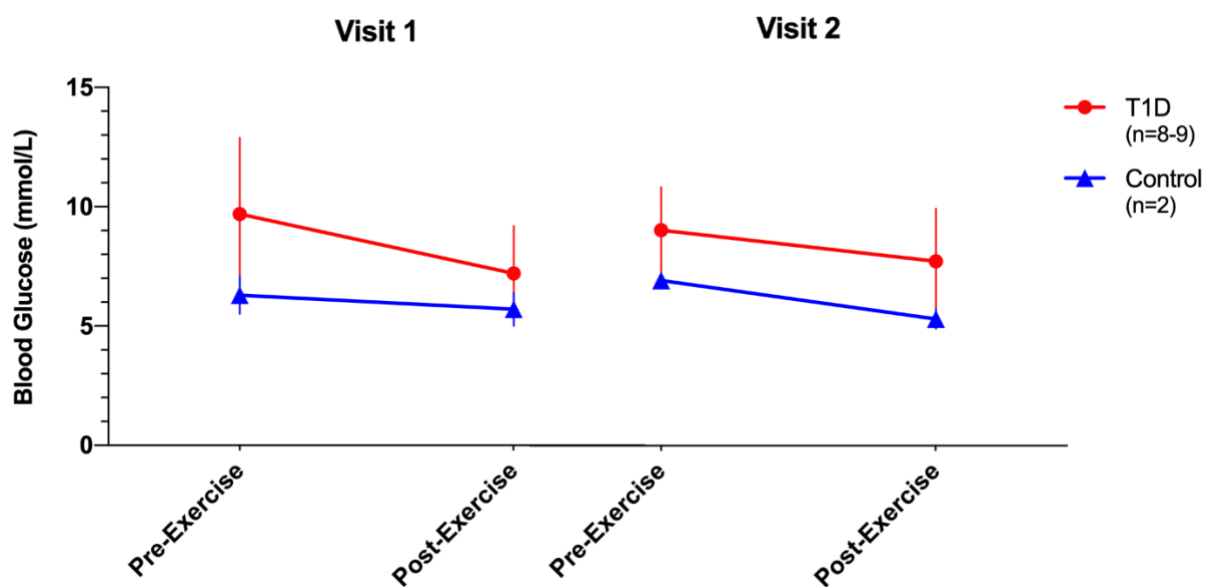


Figure 15. Blood Glucose. Blood glucose (BG) from pre- to post-exercise in T1D and control subjects during visit 1 and visit 2 in Study B. BG measurements were taken using SMBG.

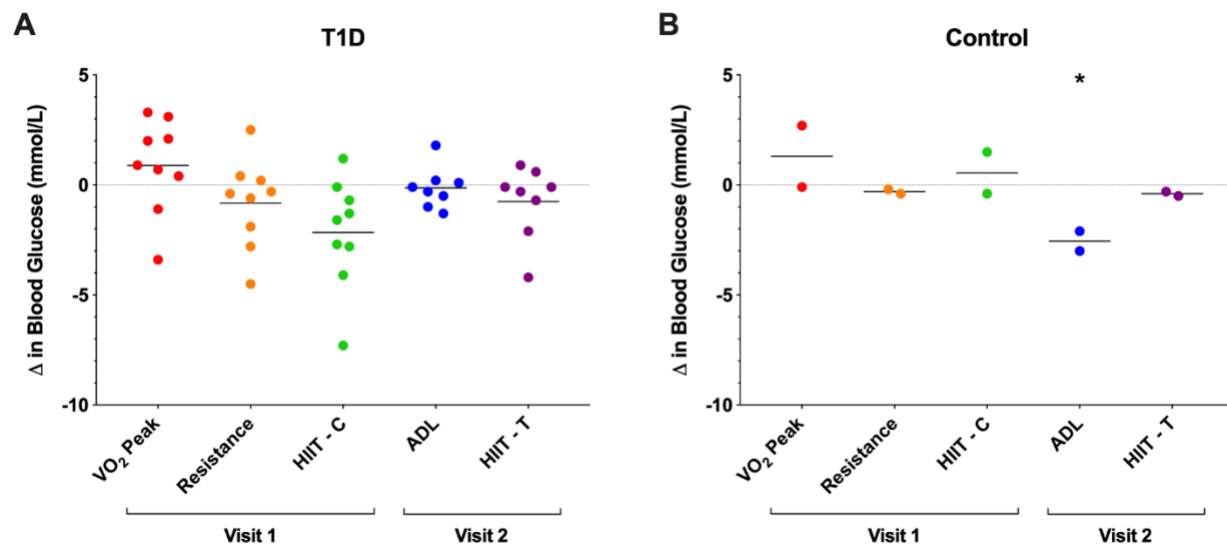


Figure 16. Change in Blood Glucose. Change in blood glucose in T1D (A) and control subjects (B) from the start to the end of the VO₂ peak test, resistance exercise, activities of daily living (ADL), and high-intensity interval training on a treadmill (HIIT-T) and cycle ergometer (HIIT-C) in Study B. BG measurements were taken using SMBG. The black horizontal lines represent the mean. n=2-9. * $P < 0.05$ compared to T1D.

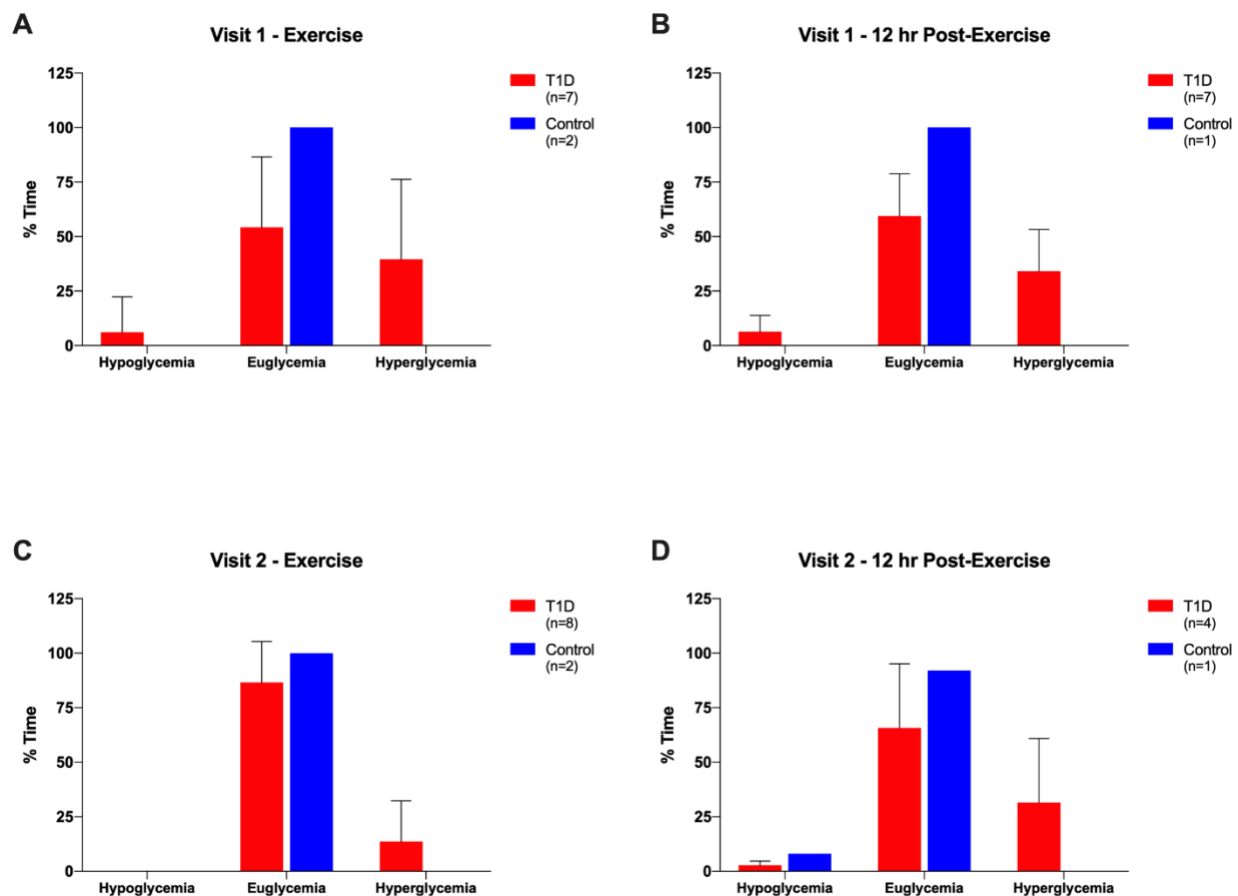


Figure 17. Percent Time in Hypoglycemia, Euglycemia, and Hyperglycemia. Percent time spent in hypoglycemia, euglycemia, and hyperglycemia during exercise (A) and for 12 hours post-exercise (B) in visit 1 and during exercise (C) and for 12 hours post-exercise (D) in visit 2 by T1D and control subjects in Study B. Glucose measurements were taken using CGM. All data are mean \pm SD.

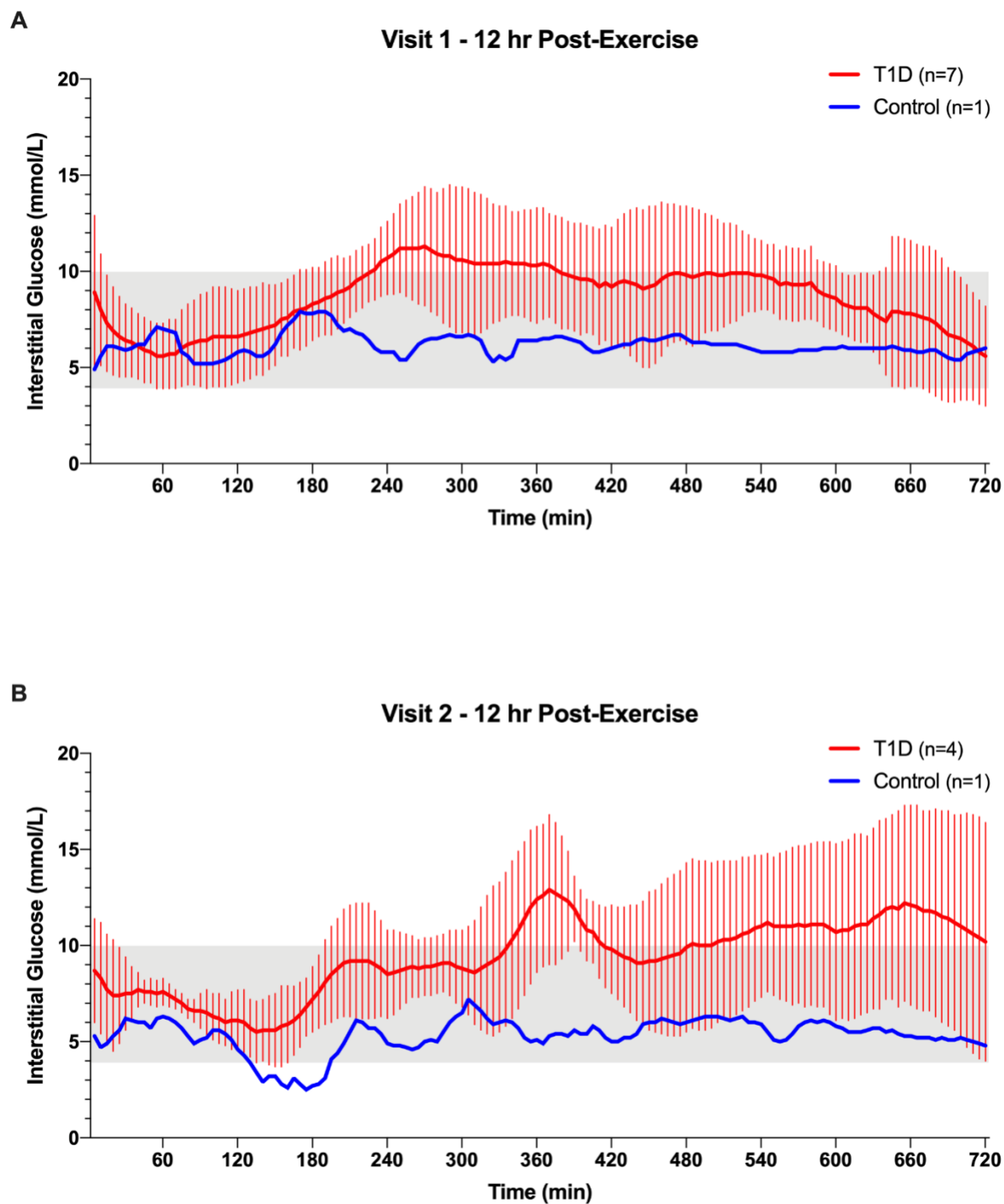


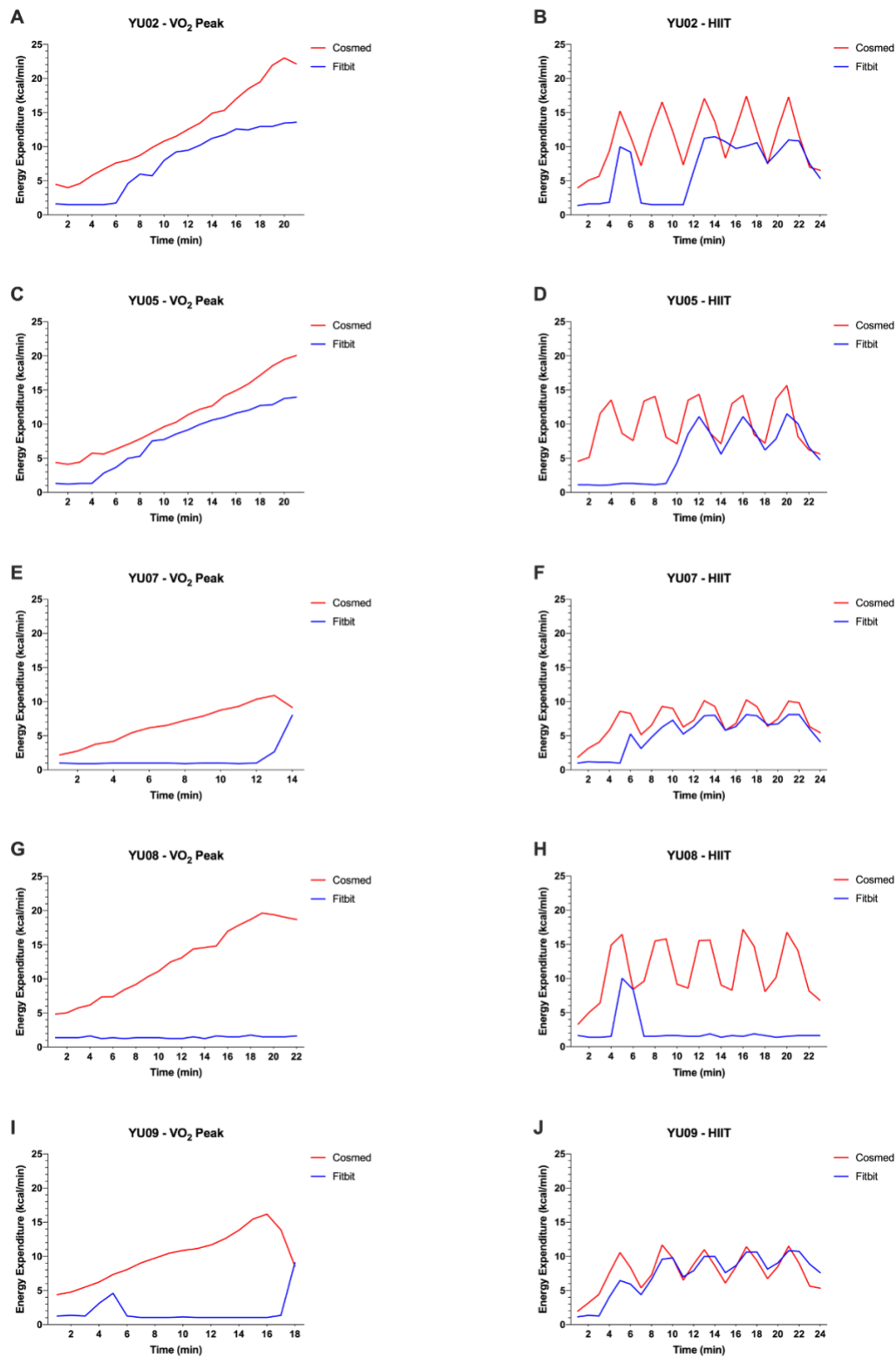
Figure 18. Interstitial Glucose. Interstitial glucose tracings from CGM for 12 hours post-exercise following visit 1 (A) and visit 2 (B) in T1D subjects and 1 control subject. The grey area represents euglycemia (3.9-10 mmol/L). All data are mean \pm SD.

10. APPENDICES

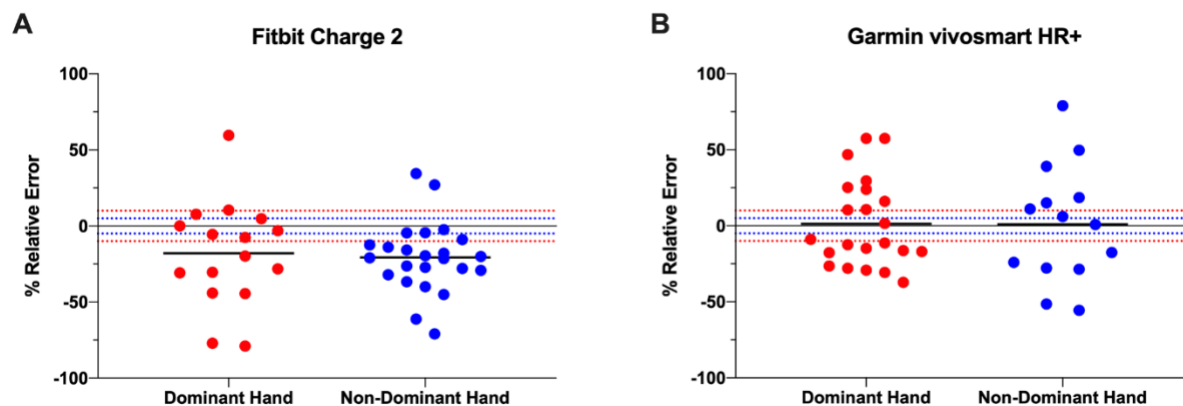
10.1. STUDY A

Appendix A. Activity Modes. Activity modes used on the Fitbit Charge 2 and Garmin vivosmart HR+ during the VO₂ peak test on a treadmill (VO₂ Peak-T) and cycle ergometer (VO₂ Peak-C), resistance exercise, activities of daily living (ADL), and high-intensity interval training on a treadmill (HIIT-T) and cycle ergometer (HIIT-C) in Study A.

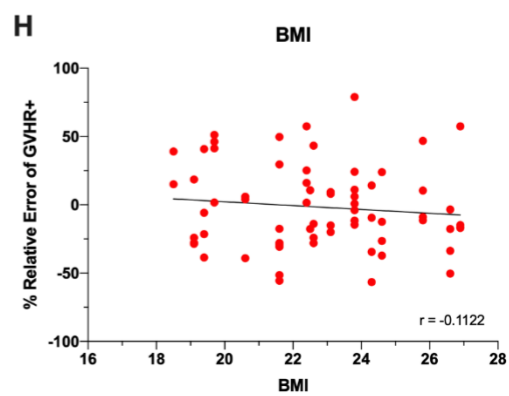
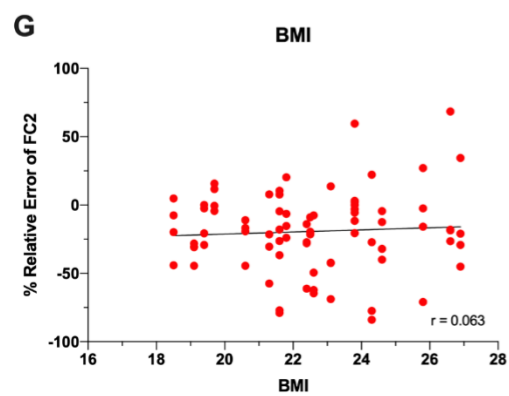
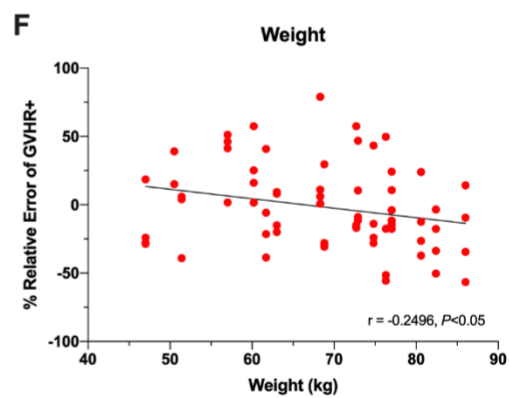
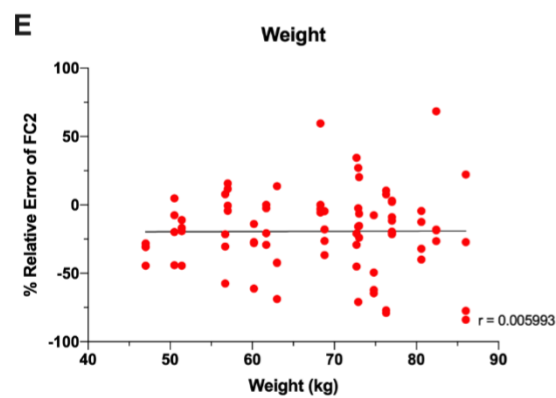
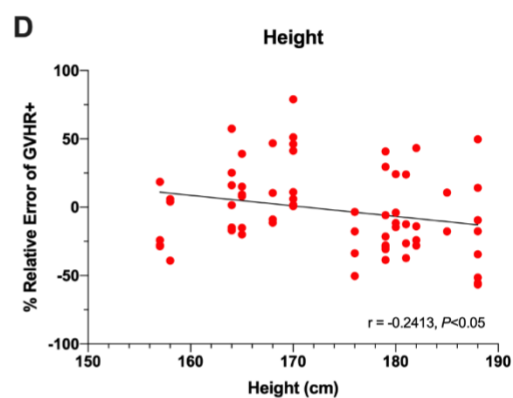
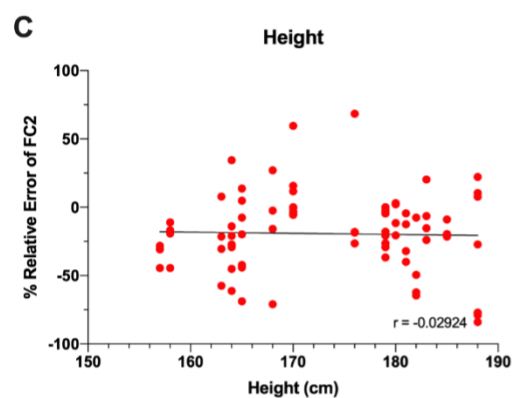
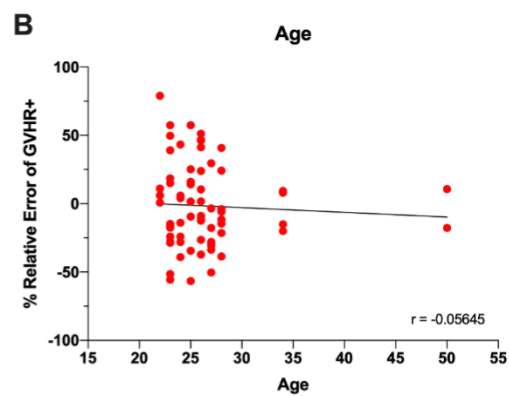
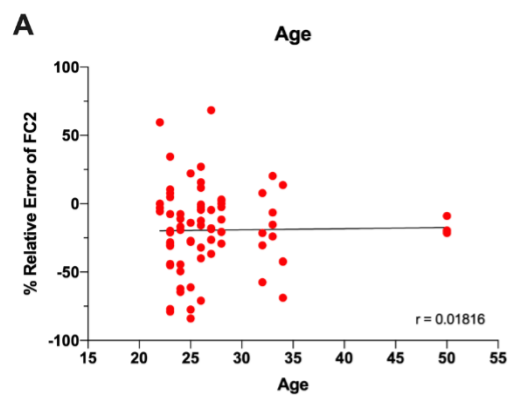
	Fitbit Charge 2	Garmin vivosmart HR+
VO ₂ Peak – T	Run	Run
VO ₂ Peak – C	Bike	Cardio
Resistance	Weights	Other
ADL	Workout	Walk
HIIT – T	Run	Run
HIIT – C	Bike	Cardio

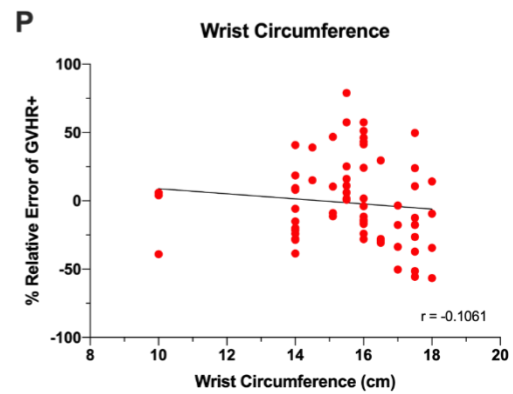
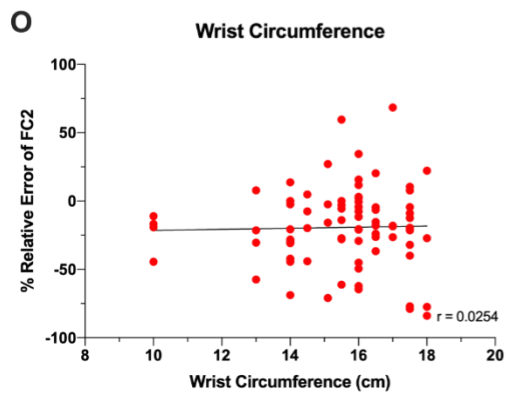
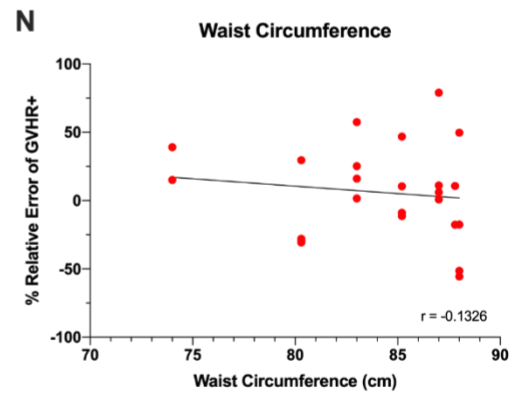
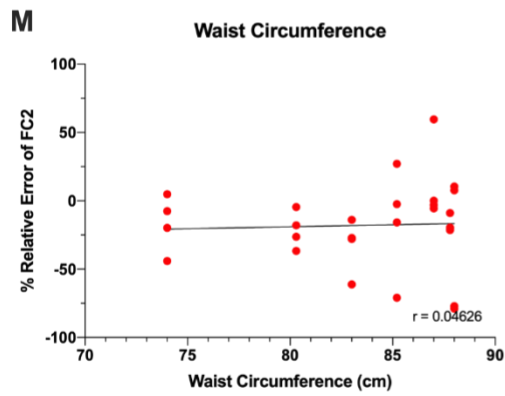
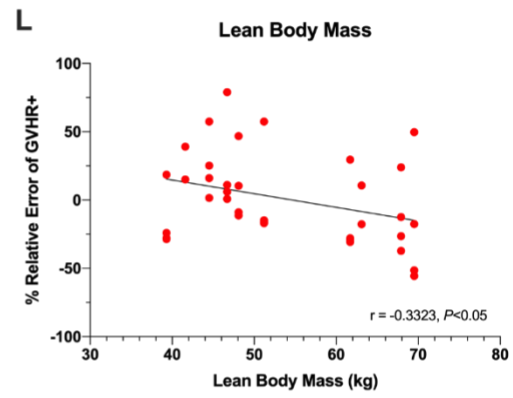
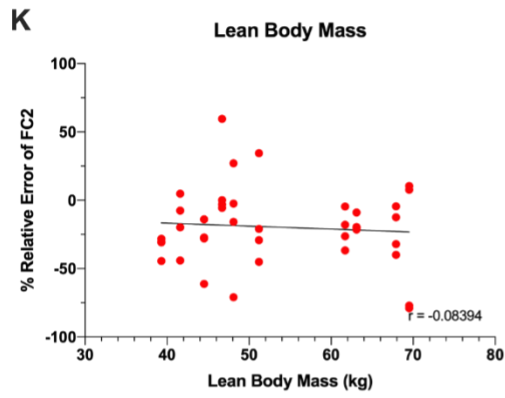
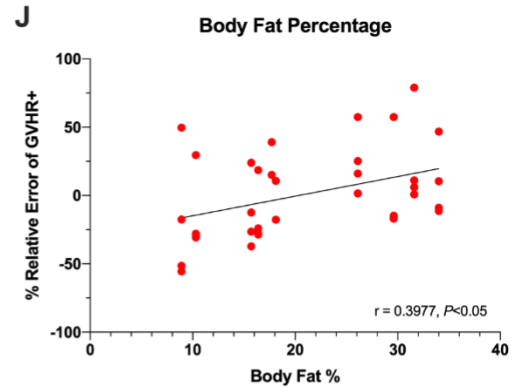
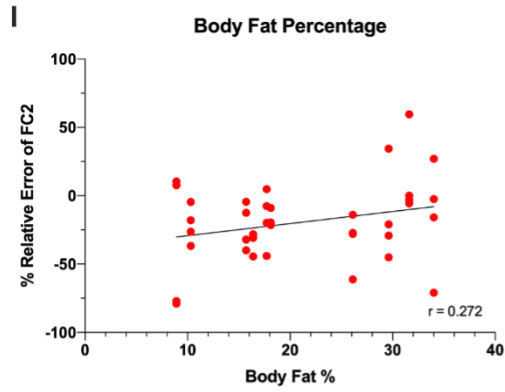


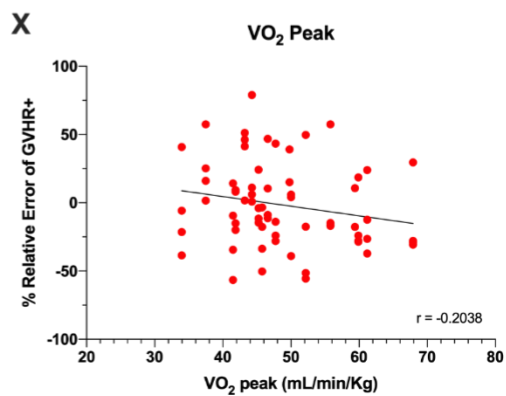
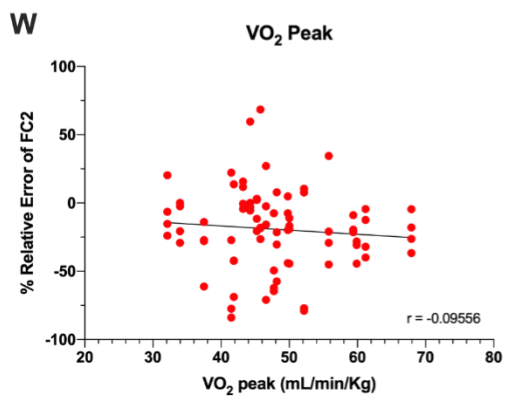
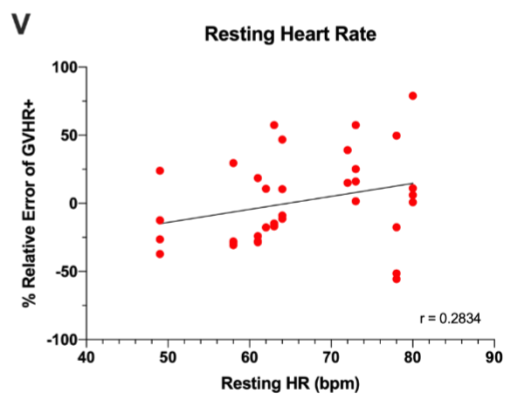
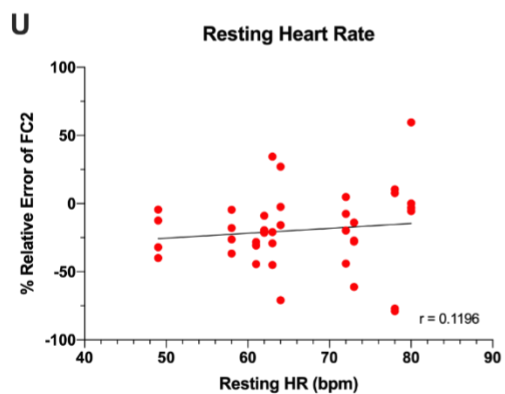
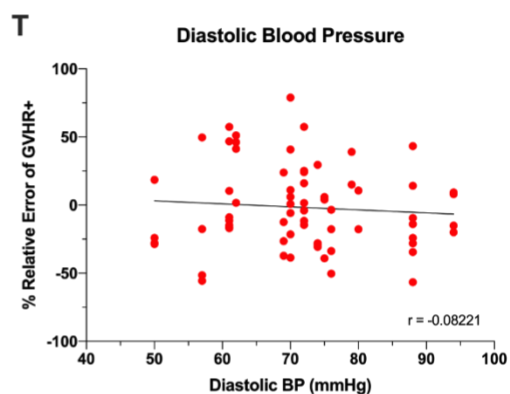
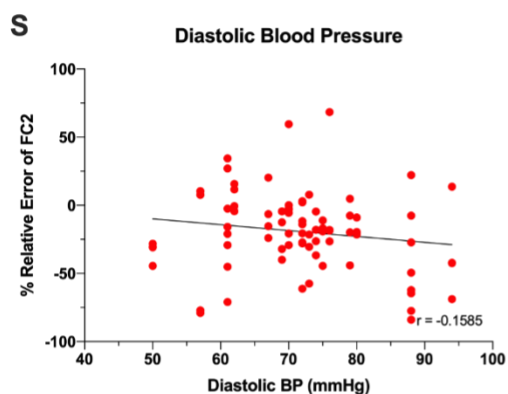
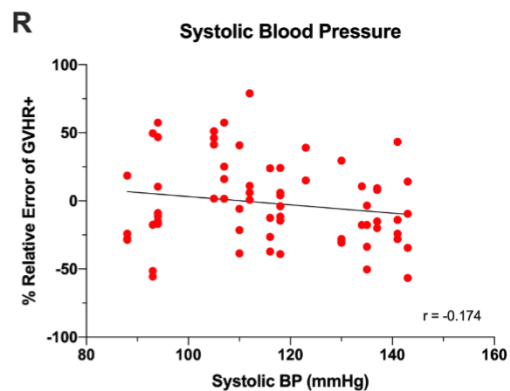
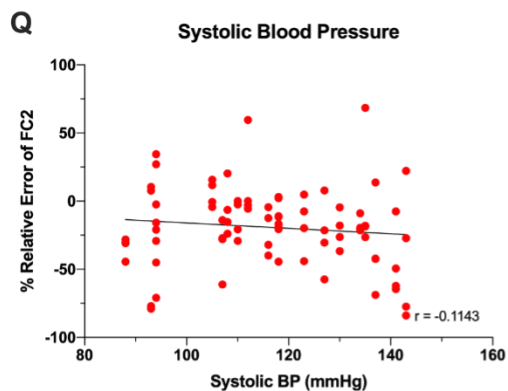
Appendix B. Energy Expenditure for Individual Subjects during Cycling Exercise. Energy expenditure (kcal/min) measurements for individual subjects (e.g. YU02) from the Fitbit Charge 2 during the VO₂ peak test (A,C,E,G,I) and high-intensity interval training (HIIT) (B,D,F,H,J) on the cycle ergometer in Study A.



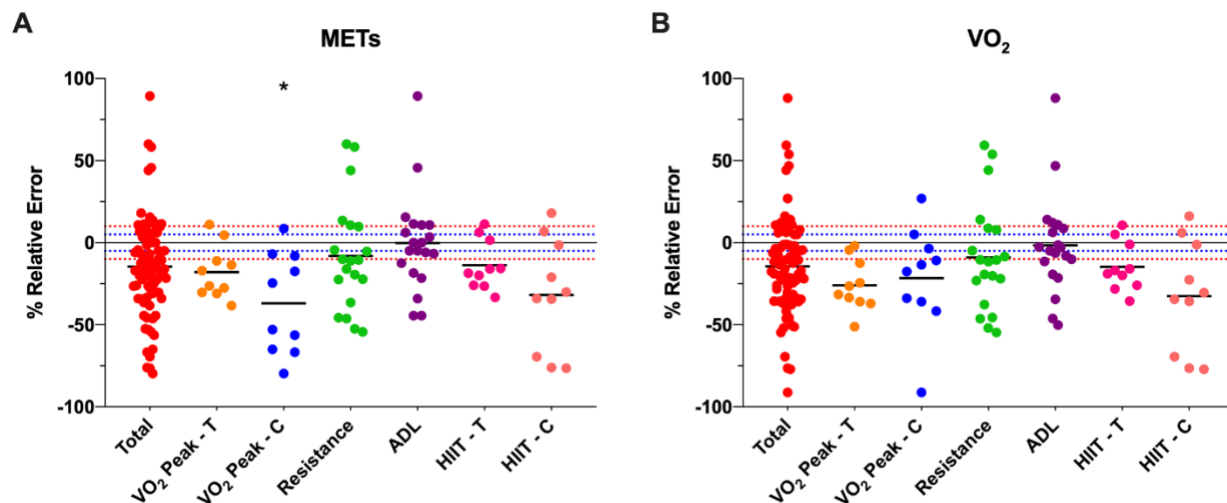
Appendix C. Relative Percent Error in Energy Expenditure for Dominant and Non-Dominant Hand Wear. Relative percent error of the Fitbit Charge 2 (A) and Garmin vivosmart HR+ (B) in measuring energy expenditure (kcal) during dominant hand and non-dominant hand wear in Study A. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. n=4-6.







Appendix D. Correlation Plots for Relative Percent Error and Anthropometrics. Correlation plots comparing relative percent error in energy expenditure measurements (kcal) from the Fitbit Charge 2 (FC2) (A,C,E,G,I,K,M,O,Q,S,U,W) and Garmin vivosmart HR+ (GVHR+) (B,D,F,H,J,L,N,P,R,T,V,X) to subject's age (A-B), height (C-D), weight (E-F), BMI (G-H), body fat percentage (I-J), lean body mass (K-L), waist circumference (M-N), wrist circumference (O-P), systolic blood pressure (Q-R), diastolic blood pressure (S-T), resting heart rate (U-V), and VO₂ peak (W-X) in Study A. The black solid line identifies linear regression. n=6-20.

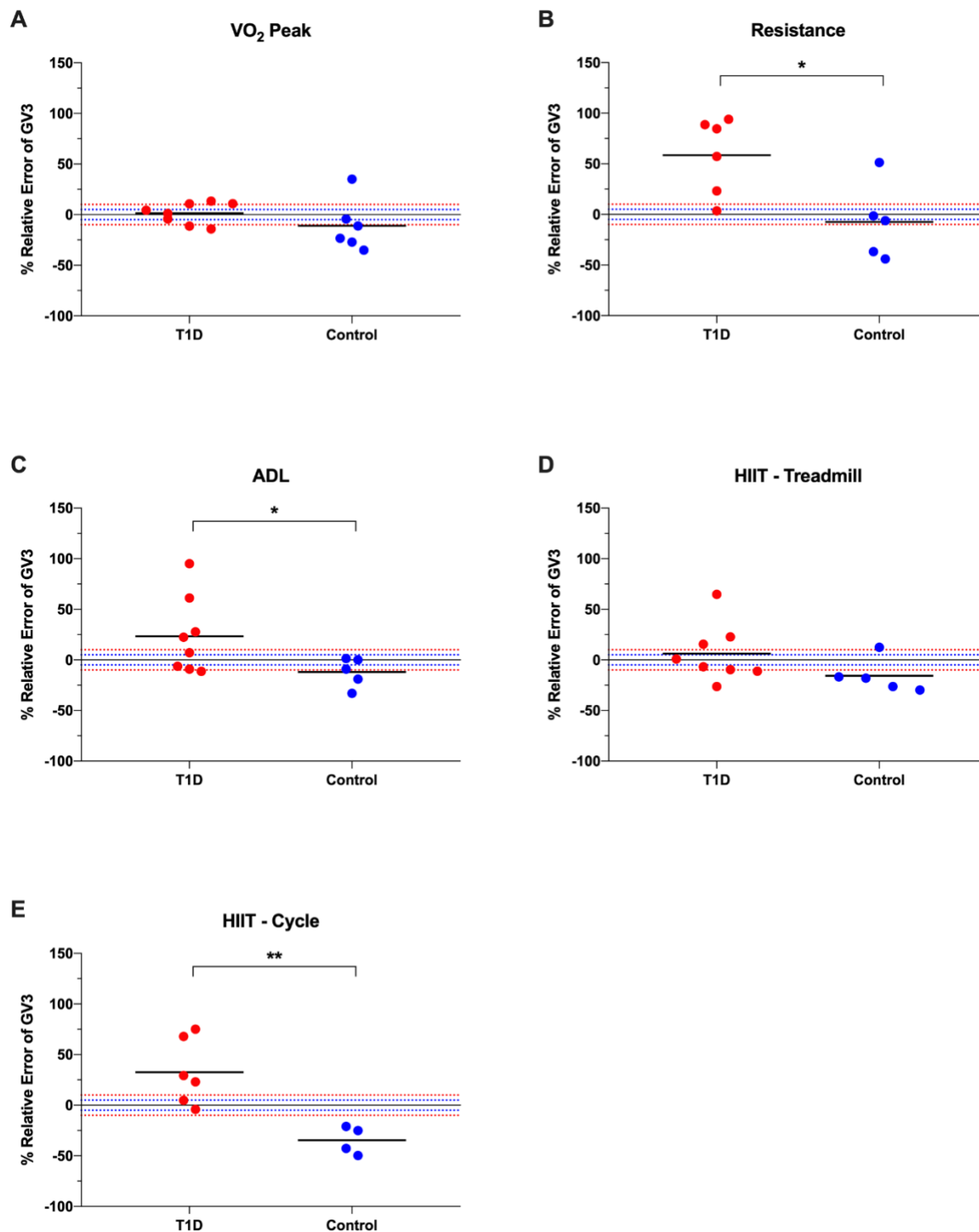


Appendix E. Relative Percent Error in METs and VO₂. Relative percent error of the Fitbit Charge 2 in measuring METs (A) and VO₂ (B) during the VO₂ peak test on a treadmill (VO₂ Peak-T) and cycle ergometer (VO₂ Peak-C), resistance exercise, activities of daily living (ADL), and high-intensity interval training on a treadmill (HIIT-T) and cycle ergometer (HIIT-C) in Study A. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. n=10-20. **P*<0.05 compared to ADL.

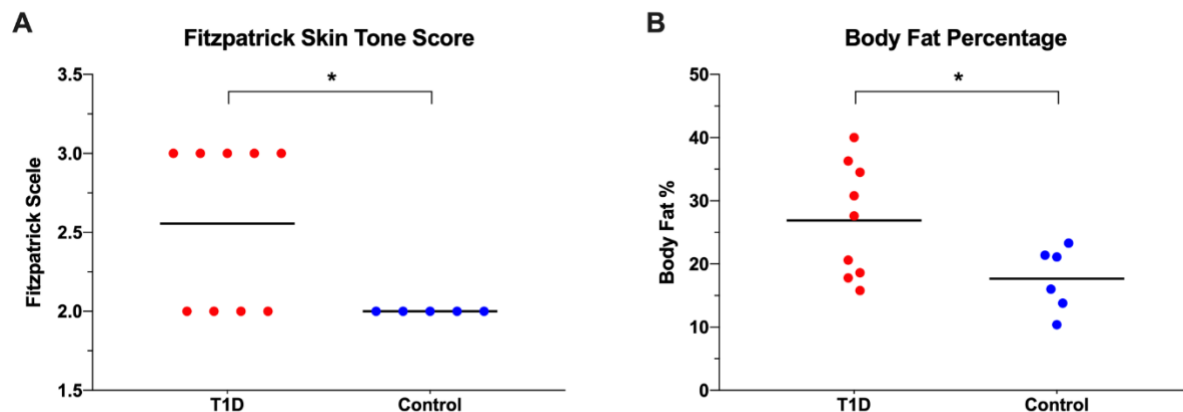
10.2. STUDY B

Appendix F. Activity Modes. Activity modes used on the Fitbit Ionic and Garmin vivosmart 3 during the VO₂ peak test, resistance exercise, activities of daily living (ADL), and high-intensity interval training on a treadmill (HIIT-T) and cycle ergometer (HIIT-C) in Study B.

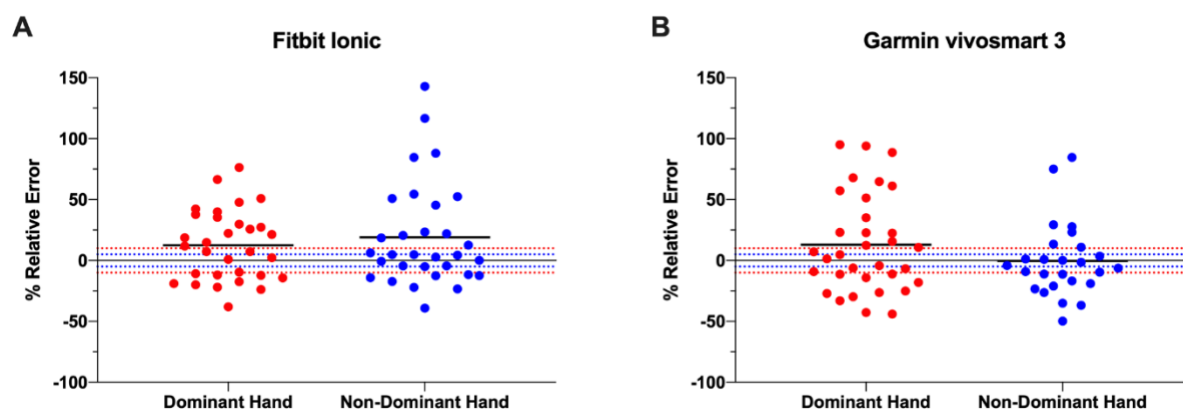
	Fitbit Ionic	Garmin vivosmart 3
VO ₂ Peak	Treadmill	Run
Resistance	Weights	Other
ADL	Workout	Walk
HIIT – T	Treadmill	Run
HIIT – C	Bike	Cardio



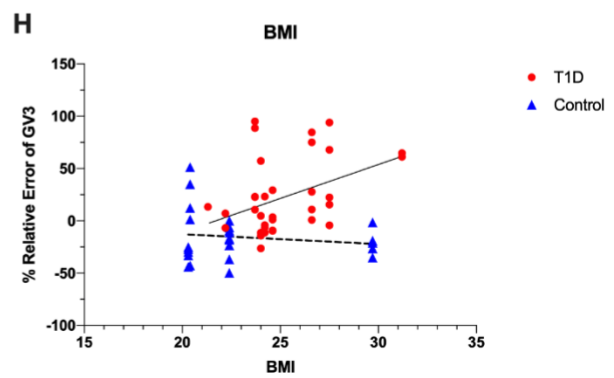
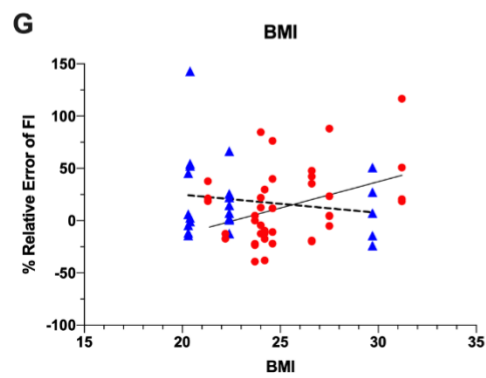
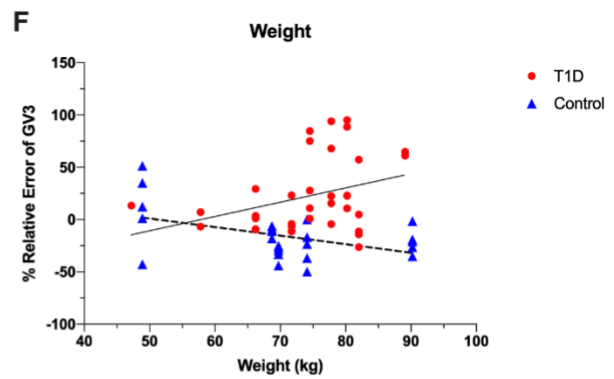
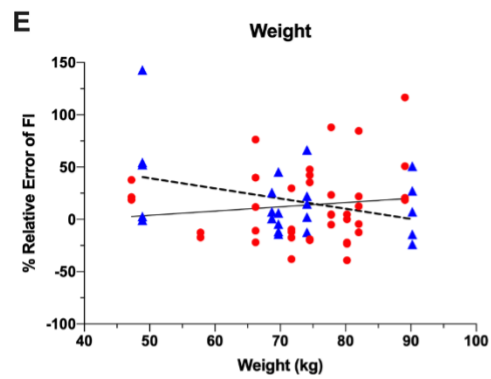
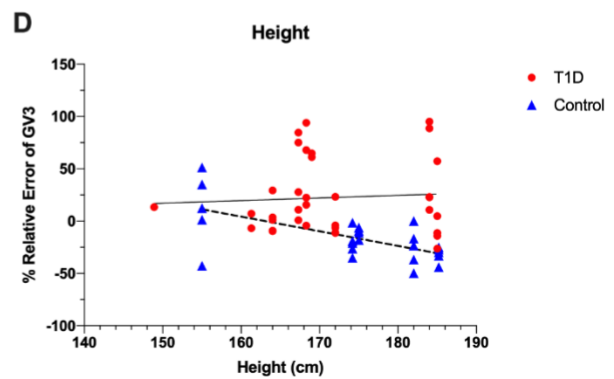
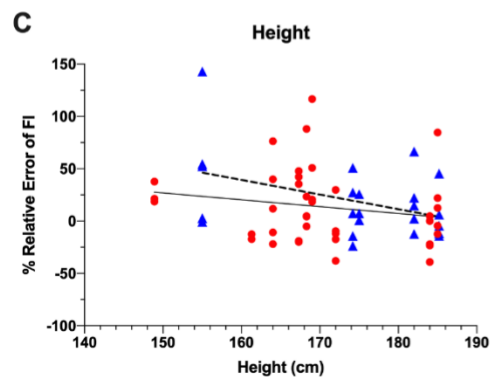
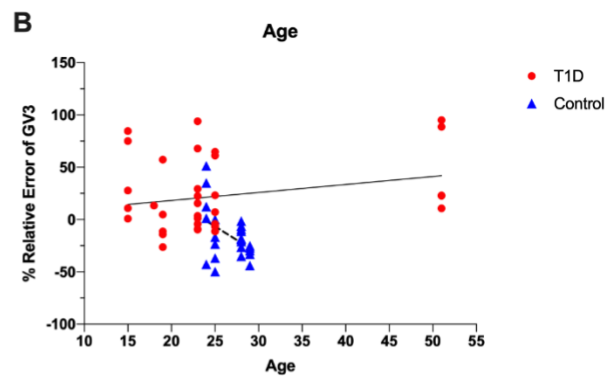
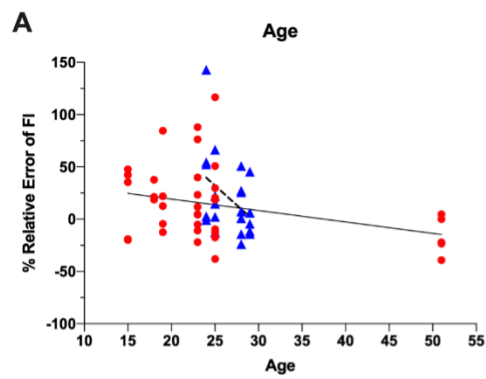
Appendix G. Relative Percent Error in Energy Expenditure on Garmin vivosmart 3. Relative percent error in measuring energy expenditure (kcal) on the Garmin vivosmart 3 (GV3) during the VO₂ peak test (A), resistance exercise (B), activities of daily living (ADL) (C), and high-intensity interval training (HIIT) on a treadmill (D) and cycle ergometer (E) in Study B. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. n=10-13. * $P < 0.05$. ** $P < 0.01$.

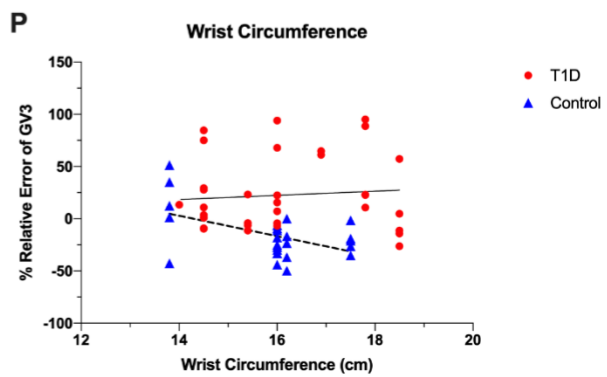
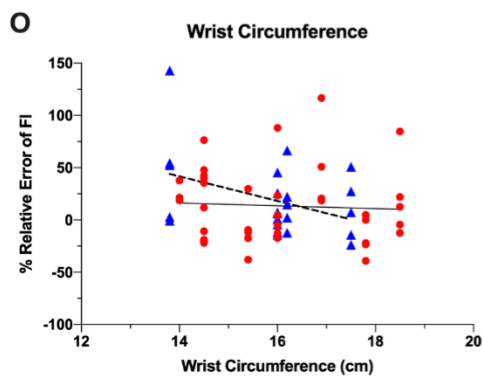
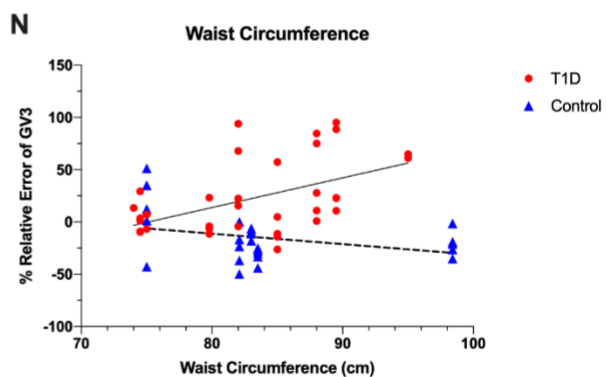
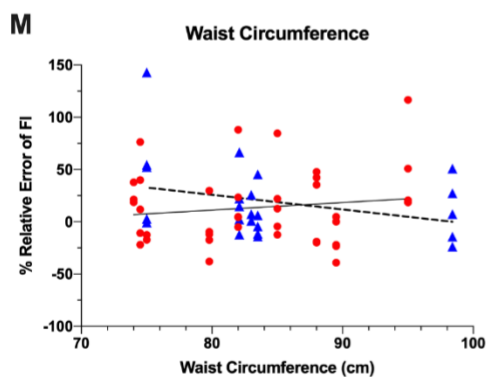
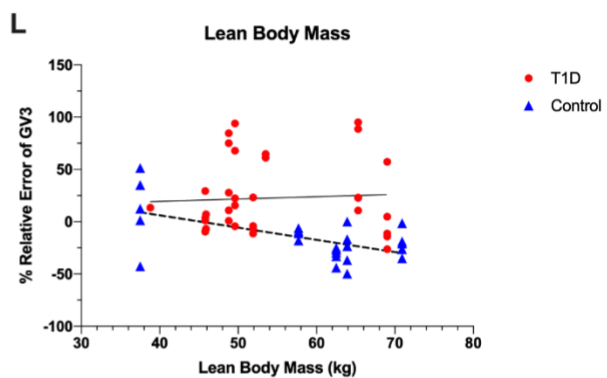
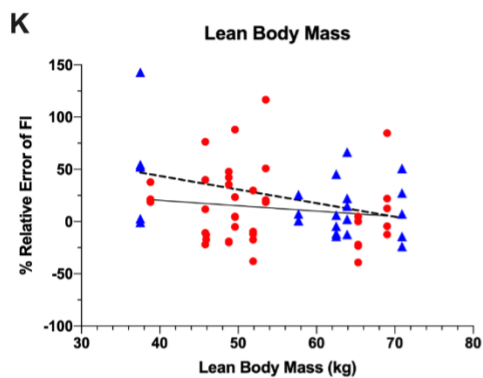
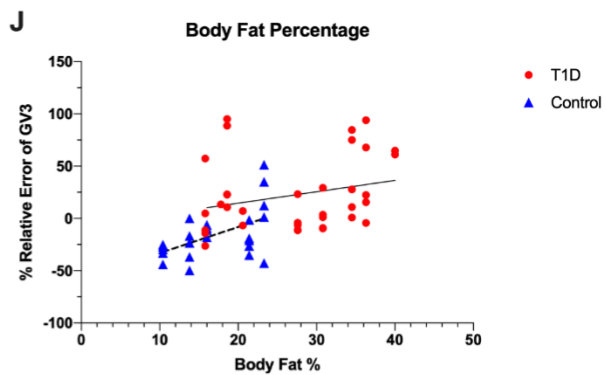
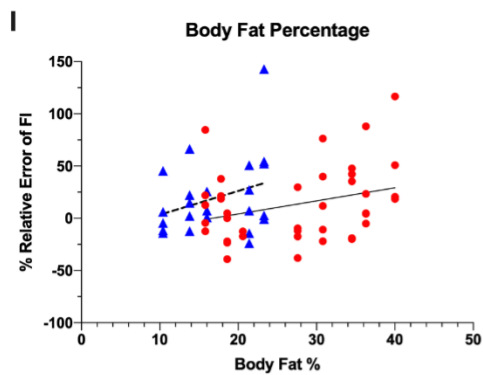


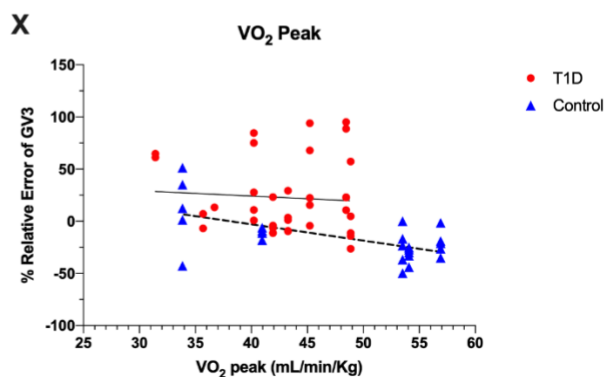
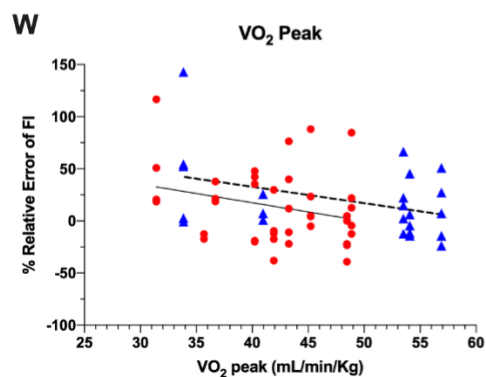
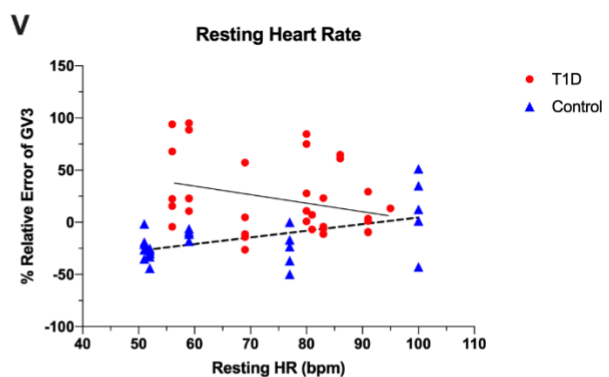
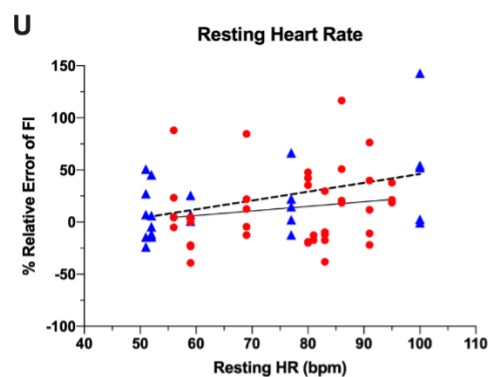
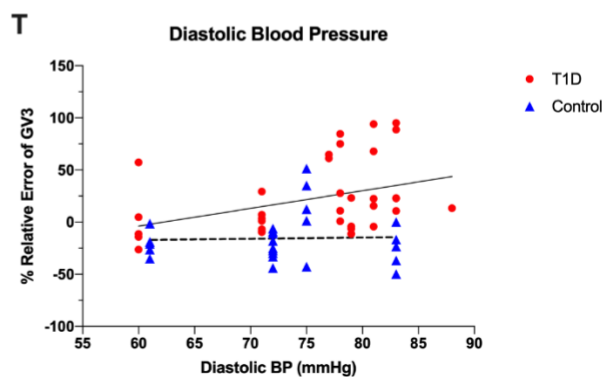
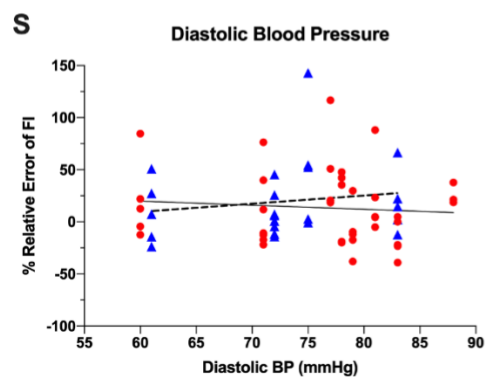
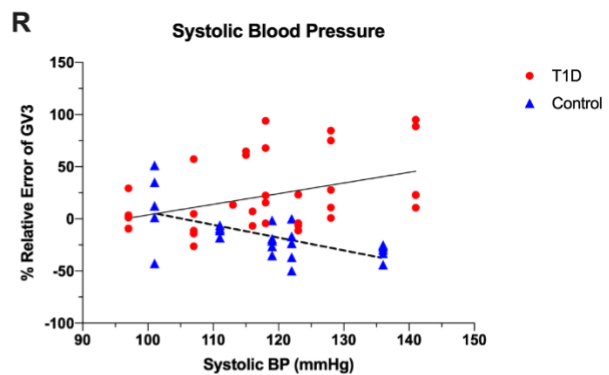
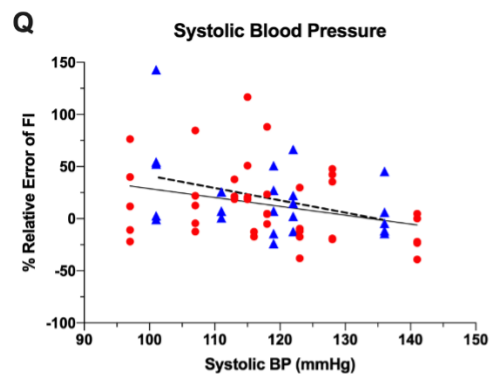
Appendix H. Fitzpatrick Skin Tone Score and Body Fat Percentage in T1D and Control Subjects. Fitzpatrick skin tone score (A) and body fat percentage (B) of T1D and control subjects in Study B. The black horizontal lines represent the mean. n=6-9. * $P < 0.05$.

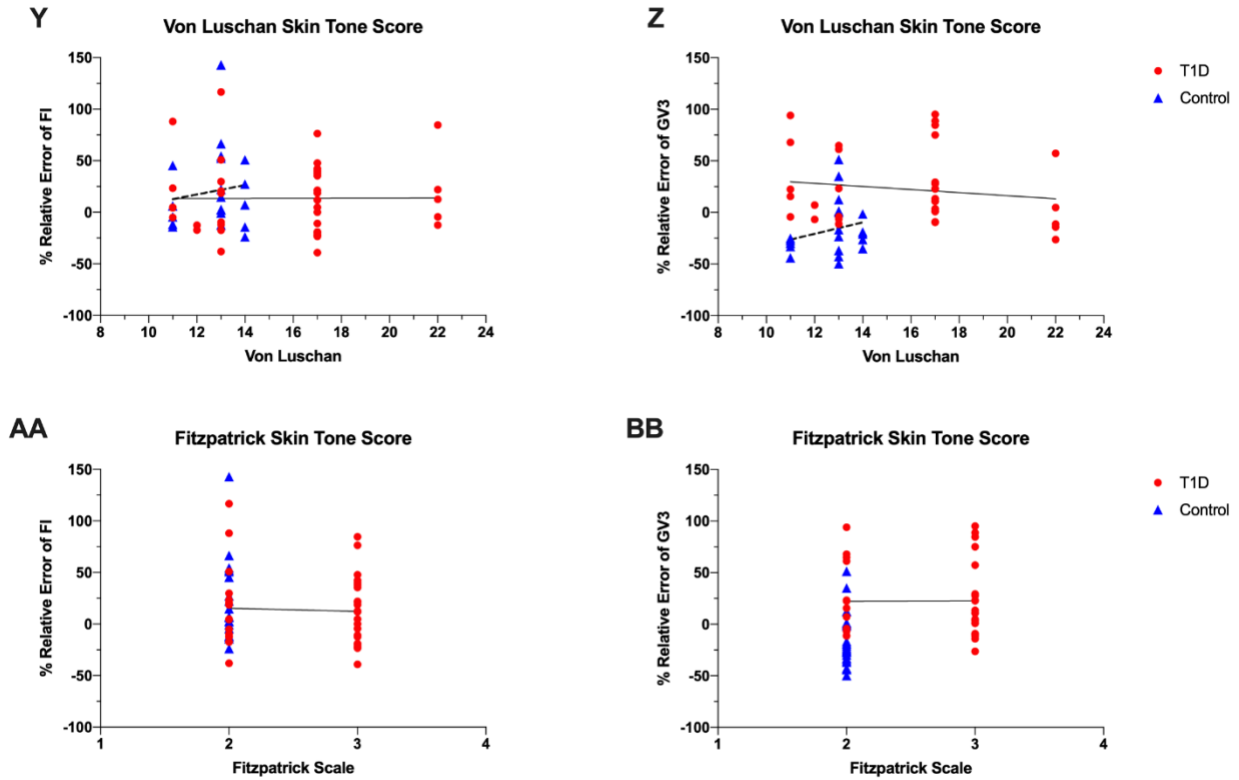


Appendix I. Relative Percent Error in Energy Expenditure for Dominant and Non-Dominant Hand Wear. Relative percent error of the Fitbit Ionic (A) and Garmin vivosmart 3 (B) in measuring energy expenditure (kcal) during dominant hand and non-dominant hand wear in Study B. The black horizontal lines represent the mean. The blue dashed lines indicate the 5% error threshold and the red dashed lines indicate the 10% error threshold. n=6-8.









Appendix J. Correlation Plots for Relative Percent Error and Anthropometrics. Correlation plots comparing relative percent error in energy expenditure measurements (kcal) from the Fitbit Ionic (FI) (A,C,E,G,I,K,M,O,Q,S,U,W,Y,AA) and Garmin vivosmart 3 (GV3) (B,D,F,H,J,L,N,P,R,T,V,X,Z,BB) to subject's age (A-B), height (C-D), weight (E-F), BMI (G-H), body fat percentage (I-J), lean body mass (K-L), waist circumference (M-N), wrist circumference (O-P), systolic blood pressure (Q-R), diastolic blood pressure (S-T), resting heart rate (U-V), VO_2 peak (W-X), Von Lushan skin tone score (Y-Z), and Fitzpatrick skin tone score (AA-BB) in Study B. The black solid line identifies linear regression for T1D subject data and the black dashed line identifies linear regression for control subject data. $n=10-13$.